

# WARSAW UNIVERSITY OF TECHNOLOGY

**Faculty of Electronics  
and Information Technology**

## DOCTORAL DISSERTATION

mgr Piotr Szczepański

**Fast Algorithms for Game-Theoretic Centrality Measures**

Supervisor  
Prof. Mieczysław Muraszekiewicz

Secondary Supervisor  
PhD. Tomasz Michalak

Warsaw 2015



*For my wonderful wife, Paula.*



## Abstract

In this dissertation, we analyze the computational properties of game-theoretic centrality measures. The key idea behind game-theoretic approach to network analysis is to treat nodes as players in a cooperative game, where the value of each coalition of nodes is determined by certain graph properties. Next, the centrality of any individual node is determined by a chosen game-theoretic solution concept (notably, the Shapley value) in the same way as the payoff of a player in a cooperative game. On one hand, the advantage of game-theoretic centrality measures is that nodes are ranked not only according to their individual roles but also according to how they contribute to the role played by all possible subsets of nodes. On the other hand, the disadvantage is that the game-theoretic solution concepts are typically computationally challenging. The main contribution of this dissertation is that we show that a wide variety of game-theoretic solution concepts on networks can be computed in polynomial time.

Our focus is on centralities based on the Shapley value and its various extensions, such as the Semivalues and Coalitional Semivalues. Furthermore, we prove #P-hardness of computing the Shapley value in connectivity games and propose an algorithm to compute it. Finally, we analyse computational properties of generalized version of cooperative games in which order of player matters. We propose a new representation for such games, called generalized marginal contribution networks, that allows for polynomial computation in the size of the representation of two dedicated extensions of the Shapley value to this class of games.

**Keywords:** *the Shapley value, the Owen value, Semivalues, social networks, centrality measures*

## Streszczenie

W niniejszej rozprawie autor porusza problem złożoności obliczeniowej teoriogrowych centralności. W teoriogrowym podejściu do analizy sieci traktujemy wierzchołki jako graczy w koalicyjnej grze, w której wartość koalicji wierzchołków wynika ze struktury sieci. W takiej grze ważność każdego wierzchołka może być określona przez dowolne rozwiązanie gry koalicyjnej (w szczególności przez wartość Shapleya). Z jednej strony zaletą takiego podejścia do centralności jest fakt, że ranking wierzchołków wynika nie tylko z indywidualnej roli każdego wierzchołka, lecz także z tego, ile dany wierzchołek wnosi do roli każdego podzbioru wierzchołków w grafie. Jednakże z drugiej strony liczenie rozwiązań gier koalicyjnych jest bardzo trudne. Główną kontrybucją tej rozprawy jest pokazanie, jak liczyć w czasie wielomianowym teoriogrowe centralności dla wielu różnych rozwiązań gier koalicyjnych.

W tej pracy autor skupia się na szybkim liczeniu teoriogrowych centralności opartych na wartości Shapleya i jej różnych rozszerzeniach. W szczególności analizuje on obliczeniowe własności Półwartości, będące uogólnieniem wartości Shapleya, i koalicje Półwartości, będące uogólnieniem wartości Owena. Dodatkowo, autor dowodzi #P-trudności liczenia wartości Shapleya w grach spójnościowych i proponuje szybki algorytm radzący sobie z tym problemem. Na zakończenie, w pracy analizowane są uogólnione gry koalicyjne, w których kolenność graczy wyznacza wartość koalicji. Autor proponuje nową, zwięzłą reprezentację tych gier, która pozwala na obliczanie w czasie wielomianowym ze względu na wielkość reprezentacji dwóch dedykowanych tym grom rozszerzeń wartości Shapleya.

**Słowa kluczowe:** *wartość Shapleya, wartość Owena, Półwartości, sieci społecznościowe, miary centralności*

# Contents

<b>1. Introduction</b>	11
1.1. Publications and Author's Contribution	13
<b>2. Preliminaries</b>	16
2.1. Coalitional Games	16
2.2. The Shapley Value	18
2.3. Semivalues	21
2.4. Marginal Contribution Networks	22
2.5. Read-once Marginal Contribution Networks	25
2.6. Graph Theory	26
2.7. Centrality Measures	26
2.8. Community Structures	29
<b>3. Related Work</b>	30
3.1. Centrality Measures and Their Classification	30
3.2. Game-theoretic Approach to Centrality Measures	31
3.3. Computational Aspects of Coalitional Games – Representations	33
<b>4. Polynomial Algorithms for Game-theoretic Centrality Measures</b>	36
4.1. Why Use Game-Theoretic Centrality Measures?	37
4.2. General Definition of Game-Theoretic Network Centrality	39
4.3. The Algorithms	42
4.3.1. Shapley Value-Based Degree Centrality	43
4.3.2. Shapley Value-Based $k$ -influence Degree Centrality	45
4.3.3. Shapley Value-Based $d_{cutoff}$ -distance Degree Centrality	47
4.3.4. Shapley Value-Based Closeness Centrality	49
4.3.5. Shapley Value-Based $W_{cutoff}$ -influence Degree Centrality	52
4.4. Simulations	53
4.4.1. Approximation Methods for the Shapley Value	54
4.4.2. Simulation Setup	55
4.4.3. Data Used in Simulations	59
4.4.4. Simulation Results	59
4.5. Conclusions	62
<b>5. Efficient Algorithms for a Game-Theoretic Measure of Betweenness Centrality</b>	65

5.1.	Introduction . . . . .	65
5.2.	The Overview of Computational Results . . . . .	68
5.3.	The New Centrality and Its Properties . . . . .	69
5.3.1.	Motivating Scenarios . . . . .	69
5.3.2.	The Shapley Value-based Betweenness Centrality . . . . .	72
5.3.3.	Semivalue-based Betweenness Centrality . . . . .	75
5.3.4.	A Look at Marginal Contribution for the Shapley Value . . . . .	76
5.3.5.	A Look at Marginal Contribution for the Semivalue . . . . .	79
5.4.	Algorithms to Compute the Shapley Value and Semivalue based Betweenness Centralities	81
5.4.1.	The Framework for Unweighted Graphs . . . . .	82
5.4.2.	The Algorithm for the Shapley Value-based Betweenness Centrality for Unweighted Graph . . . . .	84
5.4.3.	The Algorithm for the Semivalue-based Betweenness Centrality for Unweighted Graphs . . . . .	85
5.4.4.	The Framework for Weighted Graphs . . . . .	86
5.4.5.	The Algorithm for Shapley Value-based Betweenness Centrality in Weighted Graphs . . . . .	88
5.4.6.	The Algorithm for Semivalue-based Betweenness Centrality in Weighted Graphs	90
5.5.	Simulations . . . . .	91
5.5.1.	Evaluating our Centrality Measure . . . . .	91
5.5.2.	Evaluating our Algorithms . . . . .	96
5.6.	Conclusions . . . . .	99
<b>6.</b>	<b>A Centrality Measure for Networks With Community Structure Based on a Generalization of the Owen Value . . . . .</b>	<b>101</b>
6.1.	Introduction . . . . .	101
6.2.	Preliminaries . . . . .	102
6.2.1.	The Coalitional Structure . . . . .	102
6.2.2.	The Owen Value . . . . .	103
6.3.	The Coalitional Semivalue . . . . .	104
6.4.	The New Centrality Measure and Its Properties . . . . .	105
6.5.	Computational Analysis . . . . .	106
6.5.1.	Weighted Degree Centrality Measure . . . . .	107
6.5.2.	The Marginal Contribution Analysis . . . . .	107
6.5.3.	Algorithms . . . . .	110
6.6.	Simulations . . . . .	111
6.7.	Conclusions . . . . .	113
<b>7.</b>	<b>Computational Analysis of a Connectivity Game on Networks . . . . .</b>	<b>114</b>
7.1.	Introduction . . . . .	114
7.2.	Related Work . . . . .	116



7.3.	Connectivity Games for Terrorist Networks . . . . .	117
7.4.	Computational Analysis & Algorithms . . . . .	119
7.4.1.	Complexity . . . . .	119
7.4.2.	Analysis of Marginal Contributions . . . . .	123
7.4.3.	The Basic Algorithm for Connectivity Games . . . . .	124
7.4.4.	The Faster Algorithm for Connectivity Games . . . . .	125
7.4.5.	Approximation algorithm for connectivity games . . . . .	126
7.5.	Performance Evaluation . . . . .	129
7.6.	Conclusions . . . . .	133
<b>8.</b>	<b>The Marginal Contribution Networks for Generalized Coalitional Games . . . . .</b>	<b>134</b>
8.1.	Preliminaries . . . . .	134
8.1.1.	The Generalized Coalitional Games . . . . .	134
8.1.2.	The Nowak and Radzik Value . . . . .	135
8.1.3.	The Sánchez and Bergantiños Value . . . . .	136
8.2.	Representation . . . . .	137
8.2.1.	Generalized Read-Once MC-Nets . . . . .	138
8.2.2.	Computing the Nowak-Radzik Value with Generalized Read-Once MC-Nets . . . . .	141
8.2.3.	Computing the Sánchez-Bergantiños Value with Generalized Read-Once MC-Nets . . . . .	145
8.3.	Computing Generalized Game-Theoretic Betweenness Centrality . . . . .	149
8.4.	Conclusions . . . . .	151
<b>9.</b>	<b>Summary and Potential Extensions . . . . .</b>	<b>152</b>
	<b>Appendix A. Main Notation Used in the Dissertation . . . . .</b>	<b>155</b>
	<b>Appendix B. The Recursive Functions . . . . .</b>	<b>157</b>
	<b>Bibliography . . . . .</b>	<b>159</b>



# 1. Introduction

*The Shapley* value is arguably the most well-known normative payoff division scheme (solution concept) in cooperative game theory [Shapley, 1953]. One of its interesting applications is to measure importance (or centrality) of nodes in networks. Unfortunately, although such game-theoretic approach to centrality is often more advantageous than the classical measures, the Shapley value as well as other, related solution concepts from cooperative game-theory are computationally challenging. In particular, they typically require to consider all marginal contributions that  $n$  players in the cooperative game could make to all  $2^n$  coalitions (subsets of players). Clearly, applying such a direct approach to compute game-theoretic solution concepts on networks would be prohibitive even for very small systems.

The main thesis of this dissertation is that it is possible to compute various measures based on game-theoretic solution concepts in polynomial time. This includes, among others, game-theoretic extensions of three most well-known classical centrality measures: degree, closeness and betweenness. Polynomial running time in our algorithms is achieved by probabilistic analysis of marginal contributions of players to coalitions and taking advantage of the network topology.

Networks underpin the number of real-life domains of our everyday live: from the quite obvious examples of computer networks, communications networks, or road networks to the more hidden networks like protein networks, internet network, or social networks. The studies of the structure of real-life networks were undertaken by many scientists resulting with a number of astonishing breakthroughs. Discovered and explained by Albert et al. [2000] the scale-free nature of many real-life networks were undoubtedly one of those. The other was done by Watts and Strogatz [1998] who proposed the model inspired by the network phenomenon of the *Six Degrees of Separation*. Another significant breakthrough was done much earlier by Freeman [1979] who provided the conceptual classification of centrality measures—functions evaluating the importance of nodes in the network. More specifically, Freeman distinguished three basic concepts upon which various centrality measures are built: the distance between nodes (*closeness centralities*), the number of shortest paths between nodes (*betweenness centralities*), and direct connections to other nodes (*degree centralities*).

Due to a wide variety of applications of centrality measures, analyzing and developing them is one of the major research topics in the network science. For instance, identifying key nodes in network can be used to study network vulnerability [Holme et al., 2002], to

build recommendation systems [Liu et al., 2013], to identify the focal hubs in a road network [Schultes and Sanders, 2007], to point the most critical functional entities in a protein network [Jeong et al., 2001], or to find the most influential people in a social network [Kempe et al., 2003].

The common feature of the standard centrality measures introduced by Freeman [1979] is that they assess the importance of a node by focusing only on the role that a node plays by itself. However, in many applications such an approach is inadequate due to existence *synergies* that may occur if the functioning of nodes is considered in groups.<sup>1</sup> In order to capture such synergies Everett and Borgatti [1999] introduced the concept of group centrality. Its idea is broadly the same as the one of standard centrality, but now the focus is on the functioning of a given group of nodes, rather than individual nodes. In particular, Everett and Borgatti [1999] extended three Freeman’s measures to *group degree centrality*, *group closeness centrality*, and *group betweenness centrality*.

However, while the concept of group centrality addresses the issue of quantifying synergy among nodes in a particular group, there is still another fundamental problem to be solved. In particular, it is unclear how to rank individual nodes given an exponential number of potential groups of nodes they may belong to. In other words, we need to answer the question: *how to rank individual nodes based on their group centralities?* Here, the coalitional game theory comes into play.

In particular, the coalitional (or cooperative) game theory is a part of game theory in which individual players are allowed to form coalitions with an aim to increase their profits. Now, assuming that players have all decided to cooperate (i.e. to form the *grand coalition*), one of the fundamental problems in coalitional game theory is *how to divide the payoff achieved by cooperation?* The most popular answer to this question was offered by L.S. Shapley [Shapley, 1953]—the 2012 Noble Prize Laureate—who proposed to consider *marginal contributions of players to all coalitions they could potential belong to*. He proved that there exists the unique payoff division scheme, now called the *Shapley value*, that satisfies the following four fairness axioms: *Efficiency*—the whole available payoff is distributed among players; the *Null Player*—the player that cannot contribute anything should receive zero; *Symmetry*—two players whose contributions to any coalition are always the same should be given the same payoff; and *Additivity*—the payoff division scheme should be additive.

The alternative axiomatizations of the Shapley value have been studied by numerous authors in the literature. Unfortunately, while this value has many interesting properties, computing the Shapley value is often #P-complete [Deng and Papadimitriou, 1994]. This obstacle will be overcome in this thesis in the context of game-theoretic centrality.

---

<sup>1</sup> Intuitively, synergy can be thought of as a value added from group performance. Note that synergy can be also negative and, in such a case, it is called *antergy*. For an overview of various concepts of synergy see the work by Rahwan et al. [2014].

Now we have all necessary tools to introduce game-theoretic centrality measures. In a nutshell, the idea behind them is to treat nodes as players in a coalitional game, where the value of each coalition of nodes is determined by a group centrality. In such settings, the value of each individual node can be determined by cooperative game solution such as the Shapley value. In other words, in order to *provide the ranking of individual nodes that accounts for group centrality measure*, we use the Shapley value that evaluates (fairly in a certain sense) marginal contributions of each node to all groups it could potentially belong to. The key advantage of such an approach is that nodes are ranked not only according to their individual roles in the network but also according to how they contribute to the role played by all possible subsets of nodes.

Unfortunately, as already mentioned, potential downside of game-theoretic solution concepts is that they are, in general, computationally challenging. However, in this dissertation, we show that this is not always true in the network context, i.e., we are able to compute the Shapley value and the related solution concepts in polynomial time for various centrality-related games on networks.

Our contributions in this dissertation can be summarized as follows. Firstly, we compute on networks in polynomial time the Semivalues [Dubey et al., 1981] which are the generalization of the Shapley value and offer more flexibility to define any particular game-theoretic network centrality. Secondly, we propose polynomial-time algorithm on networks for the *Owen value* [Owen, 1977] that is the most important solution concept to games with *coalitional structure*, and the *Coalitional Semivalue* our extension of the Owen value. Finally, we propose the new representation of *generalized coalitional games*, the games in which the permutations of players are considered. Our representation allows to compute Nowak-Radzik value [Nowak and Radzik, 1994] and Sánchez-Bergantiños value [Sánchez and Bergantiños, 1999]. The polynomial time algorithms for the last two values are still unknown, but we propose algorithms that significantly reduce computational complexity of the problem. We also show how to use our representation on networks.

## 1.1. Publications and Author’s Contribution

The results covered by this thesis were published in four international journals and presented on the four top conferences from Artificial Intelligence. Additionally, they gathered attention of scientific community what resulted in the number of citations.

- Chapter 4 is mostly based on the work *Efficient Computation of the Shapley Value for Game-theoretic Network Centrality* published in **Journal of Artificial Intelligence Research (JAIR)** [Michalak, Aaditha, Szczepański, Ravindran, and Jennings, 2013a]. The main two contributions presented in this chapter are: developing efficient algorithms

for computing Shapley value-based degree and closeness centralities (together with their various extensions), and providing theoretical, as well as experimental analysis of the introduced algorithms. The contribution of the author of this dissertation mostly covers the experimental part of this article. More specifically, Algorithms 1-5 were developed by Karthik Aaditha, and partially formalized by the author of this thesis. More specifically, Propositions 4.1 and 2 and proof of correctness for Algorithms 1 and 3 are the contribution of the author. Additionally, Algorithms 6-10 (for the Monte Carlo sampling of the game theoretic-centrality measures) were developed by the author of this dissertation.

- Chapter 5 is based on two publications: *A New Approach to Betweenness Centrality Based on the Shapley Value* [Szczepański, Michalak, and Rahwan, 2012] published in **Proceedings of the 11<sup>th</sup> International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)** and *Efficient Algorithms for Game-Theoretic Betweenness Centrality* published in **Artificial Intelligence Journal (AI)** [Szczepański, Michalak, and Rahwan, 2016]. All the technical content, i.e., algorithms, experiments and theorems are exclusive contributions of the author of this thesis.
- Chapter 6 is based on two publications: *A Centrality Measure for Networks With Community Structure Based on a Generalization of the Owen Value* [Szczepański, Michalak, and Wooldridge, 2014] published in **Proceedings of the 21<sup>st</sup> European Conference on Artificial Intelligence (ECAI 2014)** and *A New Approach to Measure Social Capital using Game-Theoretic Techniques* [Michalak, Moretti, Namarayan, Skibski, Szczepański, Rahwan, and Wooldridge, 2015b] published in **ACM SIGecom Exchanges**. All algorithms, experiments and theorems were developed by the author of this thesis.
- Chapter 7 is based on the publication *Computational Analysis of Connectivity Games with Application to Investigation of Terrorist Networks* [Michalak, Rahwan, Szczepański, Skibski, Narayanam, Wooldridge, and Jennings, 2013b] published in **Proceedings of the 23<sup>rd</sup> International Joint Conference on Artificial Intelligence (IJCAI 2013)**. The contribution presented in this chapter is threefold, it formally proves that computing the Shapley value in connectivity games is NP-hard, it develops faster exact algorithm for connectivity games, and it develops one approximation algorithm for these family of games. The complexity proof consists of two theorems: Theorem 4 developed by the author with help of Colin McQuillan and Theorem 5, which was developed exclusively by the author of this dissertation. Algorithm 20 is a joint work done by the author, Tomasz Michalak and Talal Rahwan, and finally Algorithm 21 was developed by the author together with Oskar Skibski.
- The final Chapter 8 is based on the second part of the publication *Implementation and Computation of a Value for Generalized Characteristic Function Games* [Michalak, Szczepański, Rahwan, Chrobak, Brânzei, Wooldridge, and Jennings, 2014] published

in journal **ACM Transactions on Economics and Computation (ACM TEAC)**. The technical content in the whole chapter is an exclusive contribution of the author.

Some parts of the above chapters and additionally Chapter 2 (including definitions and formalization) are also based on the article *Efficient Computation of Semivalues for Game-Theoretic Network Centrality* [Szczepański, Tarkowski, Michalak, Harrenstein, and Wooldridge, 2015b] published in **Proceedings of the 29<sup>th</sup> AAAI Conference on Artificial Intelligence (AAAI 2015)**.

## 2. Preliminaries

In this chapter we present the basic notation and definitions from both graph theory and game theory that will be used throughout the thesis.

Firstly, we introduce *coalitional games* (also known as *cooperative games*), where we focus on the key solution concepts and their computational aspects. In particular, we define the *Shapley value*—the most popular normative solution concept to cooperative games. We also introduce a parametrized generalization to the Shapley value called the *Semivalue*. Since the standard model (or representation) of cooperative games, i.e. the *characteristic function*, is computationally challenging, we discuss an alternative representation, called the *Marginal Contribution Networks*. For certain games, this representation allows for computing the Shapley value in polynomial time in the number of agents.

Secondly, we define the basic notation of a graph and introduce elementary concepts related to social network analysis such as: *centralities* and *communities*.

### 2.1. Coalitional Games

Game Theory consists of two broad areas: *non-cooperative* (or *strategic*) games and *cooperative* (or *coalitional*) games. The first class of games is one in which players make decisions independently. In the second class, groups of players, called *coalitions*, are allowed to form profitable coalitions [Osborne and Rubinstein, 1994].

**Definition 1 (Coalitional game).** A coalitional game consists of a set of players (or agents)  $A = \{a_1, a_2, \dots, a_{|A|}\}$ , and the characteristic function  $\nu : 2^A \rightarrow \mathbb{R}$ , which assigns to each coalition of players  $C \subseteq A$  a real value (or payoff) indicating its performance, where  $\nu(\emptyset) = 0$ . Thus, the coalitional game  $g$  in the characteristic function form is a pair  $g = (A, \nu)$ .

In such defined game the number of all coalitions equals  $2^{|A|}$ —the number of all subsets of  $A$ . The set of all coalitional games will be denoted by  $\mathcal{V}$ , so we can write  $(A, \nu) \in \mathcal{V}$ , or sometimes we simplify notation and write  $\nu \in \mathcal{V}$ . The coalition of all players  $A$  is called *grand coalition*.

**Example 1.** Let us consider the game where players collect cards with famous rock stars. Each card has its own value, which could be doubled if only player possesses all cards from the set.



The game consists of two sets of cards  $K = \{k_1, k_2\}$   $M = \{m_1, m_2, m_3\}$ , and four players  $A = \{a_1, a_2, a_3, a_4\}$  who have the following cards:

$$\begin{aligned} a_1 &\longleftrightarrow k_1, m_3 \\ a_2 &\longleftrightarrow k_2 \\ a_3 &\longleftrightarrow m_1 \\ a_4 &\longleftrightarrow m_2, m_3 \end{aligned}$$

Assuming that each card has individually value 1 the characteristic function describing this game  $g_1 = (A, \nu_1)$  is:

$$\begin{aligned} \nu_1(\emptyset) &= 0 \\ \nu_1(\{a_1\}) = \nu_1(\{a_2\}) = \nu_1(\{a_3\}) = \nu_1(\{a_4\}) &= 1 \\ \nu_1(\{a_2, a_3\}) &= 2 \\ \nu_1(\{a_1, a_3\}) = \nu_1(\{a_2, a_4\}) &= 3 \\ \nu_1(\{a_1, a_4\}) &= 4 \\ \nu_1(\{a_1, a_2\}) &= 5 \\ \nu_1(\{a_3, a_4\}) = \nu_1(\{a_1, a_2, a_3\}) &= 6 \\ \nu_1(\{a_2, a_3, a_4\}) = \nu_1(\{a_1, a_2, a_4\}) &= 7 \\ \nu_1(\{a_1, a_3, a_4\}) &= 8 \\ \nu_1(A) &= 11 \end{aligned}$$

Player  $a_3$  has one card  $m_1$ , so individually its value is  $\nu_1(\{a_3\}) = 1$ . The player  $a_4$  has two cards from the same set  $m_2, m_3$  and playing alone he can receive  $\nu_1(\{a_4\}) = 2$ . Now, if these two players cooperate, they will have all cards from the same set, what makes them more valuable. Playing together they will receive  $\nu_1(\{a_3, a_4\}) = 6$ , so they have a big incentive to cooperate.

Form the above example we see that the assumption  $\nu_1(\emptyset) = 0$  is natural, because the empty coalition without players cannot obtain any profits. Additionally, the characteristic function in this example incites players to collaborate, because joining new members to coalition can only improve its performance. This property of characteristic function is called *superadditivity* and may cause the forming of the grand coalition.

**Definition 2 (Superadditivity).** The characteristic function in the game  $(A, v)$  is superadditive if only the following expression holds:

$$\forall_{\substack{C_1, C_2 \subseteq A \\ C_1 \cap C_2 = \emptyset}} v(C_1 \cup C_2) \geq v(C_1) + v(C_2)$$

The property strongly connected to superadditivity is called *monotonicity*.

**Definition 3 (Monotonicity).** *The characteristic function in the game  $(A, v)$  is monotonic if only the following expression holds:*

$$\forall_{C_1 \subseteq C_2 \subseteq A} v(C_2) \geq v(C_1)$$

Caulier [2009] showed that for the games with non-negative characteristic function the superadditivity implies monotonicity. However, in general the implication in neither direction holds.

When the game is superadditive, or monotonic the *grand coalition*, i.e., the coalition of all the players in the game, has the highest value and, therefore, is formed. One of the fundamental questions in cooperative game theory is then how to divide the payoff of the grand coalition ( $v(A)$ ) among the players and the most popular normative solution to that problem is the Shapley value.

## 2.2. The Shapley Value

In principle, there may be an infinite number of divisions of the value of grand coalition among players, however, we are interested in those that meet certain desirable criteria. In order to present these criteria we introduce the notion of *marginal contribution*.

**Definition 4 (Marginal contribution).** *The marginal contribution of player  $a_i \in A$  to the coalition  $C \subseteq A \setminus \{a_i\}$  is:*

$$\text{MC}(C, a_i) = v(C \cup \{a_i\}) - v(C) \quad (2.1)$$

Now, let us consider four *fairness* properties that a given division scheme should meet. We denote by  $\phi_i(A, v)$  the payoff received by a player  $a_i$  in game  $(A, v)$ .

- Symmetry**    payoffs do not depend on the players' names.  
That is,  $\phi(A, \pi(v)) = \pi(\phi)(A, v)$  for every game  $v$  and permutation  $\pi \in \Pi(A)$ .
- Null Player**    players that make no contribution should receive nothing.  
In other words,  $(\forall C \subseteq A, \text{MC}(C, a_i) = 0) \Rightarrow (\phi_i(A, v) = 0)$ .
- Efficiency**    the entire payoff of each coalition should be distributed among its members.  
That is,  $\sum_{a_i \in A} \phi_i(A) = v(A)$ .
- Additivity**    given three games,  $g_1 = (A, v_1)$ ,  $g_2 = (A, v_2)$  and  $g_3 = (A, v_3)$ ,  
where  $v_1(C) = v_2(C) + v_3(C)$  for all  $C \subseteq A$ , the payoff of a player in  $g_1$   
should be the sum of his payoff in  $g_2$  and in  $g_3$ .

The *Symmetry* property implies that two players that contribute the same value to all coalitions should be given the same payoff. More formally:

$$(\forall C \subseteq A \setminus \{a_i, a_j\}, \text{MC}(C, a_i) = \text{MC}(C, a_j)) \Rightarrow (\phi_i(A, \nu) = \phi_j(A, \nu)) \quad (2.2)$$

The *Efficiency* property assures that the whole available payoff is distributed among players. The *Null Player* property indicates that player that can not contribute anything should receive value 0. The *Additivity* property says that if players anticipate in two separate games then their importance in the game consisting of these two games, is the sum of the importance in these two subgames. Interestingly, there exist only one division scheme that satisfies the above four criteria—Shapley value.

**Example 2.** Let us consider the coalitional game  $g_2 = (A, v)$ , in which three players  $A = \{a_1, a_2, a_3\}$  collect apples from trees. Each player posses a stand allowing him to collect apples from the top of the trees. Additionally, on two stands of the same height it is possible to put third stand. Now, let us introduce the function  $s : A \rightarrow \mathbb{N}$  evaluating the high of each stand that a given player have:

$$s(a_i) = \begin{cases} 20 & \text{if } i = 1 \\ 40 & \text{if } i = 2 \\ 20 & \text{if } i = 3 \end{cases}$$

The characteristic function  $\nu_2$  in this game is defined as follows:

$$\nu_2(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ 60 * 10 & \text{if } C = A \\ \max_{a_i \in C} \{s(a_i)\} * 10 & \text{otherwise.} \end{cases}$$

The players  $a_1$  and  $a_3$  contributes the same value to each coalition, so based on the **Symmetry** property they should have the same payoff. We also know that the **Efficiency** property implies that the whole gain 600 should be distributed among all players. Based on these two information we obtain two equations:

$$\begin{aligned} \phi_1 &= \phi_2 \\ \phi_1 + \phi_2 &= \phi_3 = 600 \end{aligned}$$

At this point we haven't enough information to solve the above equation. In order to do it, we need to use **Additivity** and **Null Player** properties.

The solution of the above game should satisfy the four *fairness* properties. To this end, Shapley [1953] proposed to evaluate the role of each player in the game proportionally to a weighted average marginal contribution of this player to all possible coalitions. Formally:

**Definition 5 (Shapley value).** In a coalitional game  $(A, \nu)$  the Shapley value for a player  $a_i$  is given by:

$$\phi_i^{SV}(A, \nu) = \sum_{C \subseteq A \setminus \{a_i\}} \frac{|C|!(|A| - |C| - 1)!}{|A|!} (v(C \cup \{a_i\}) - v(C)) \quad (2.3)$$

Importantly, the Noble prize winner L. S. Shapley [Shapley, 1953] showed that the vector  $\phi^{SV}$  is the only solution concept satisfying properties (1) – (4).

Furthermore, if the characteristic function is superadditive than Shapley value satisfies one additional property:

**Individual Rationality:** 
$$\sum_{a_i \in A} \phi_i^{SV}(\nu) \geq \nu(\{a_i\})$$

*Individual Rationality* indicates that each player has no incentive to play alone. If he do so, he will receive the payoff no greater than the payoff obtained during his collaboration with others.

The set of all players  $A$  will be sometimes considered as a ordered list ( $|A|$ -tuple)  $A = (a_1, a_2, \dots, a_{|A|})$  in order to use the notion of the permutation of all players  $\pi \in \Pi(A)$ .

The important fact about the Shapley value is that for a given player  $a_i \in A$  in the game  $(A, \nu)$  it is the expected marginal contribution of this player to the set of players  $P_\pi(a_i)$  that precede  $a_i$  in a random permutation  $\pi \in \Pi(A)$  of all players in the game.

**Definition 6 (Shapley value as an expected value).** In a coalitional game  $(A, \nu)$  the Shapley value for a player  $a_i$  is given by:

$$\phi_i^{SV}(A, \nu) = \mathbb{E}[\text{MC}(P_\pi(a_i), a_i)] = \frac{1}{|A|!} \sum_{\pi \in \Pi(A)} \text{MC}(P_\pi(a_i), a_i) \quad (2.4)$$

The equivalence between Definition 5 and Definition 6 can be easily derived by using a simple combinatorial fact: the number of permutations of  $A$  in which the set of players occurring before  $i$ -th player equals  $P_\pi(a_i)$  is exactly  $|P_\pi(a_i)|!(|A| - 1 - |P_\pi(a_i)|)!$ .

$$\begin{aligned} \phi_i^{SV}(v) &= \frac{1}{|A|!} \left( \underbrace{\text{MC}(P_{\pi_1}(a_i), a_i) + \text{MC}(P_{\pi_2}(a_i), a_i) + \dots + \text{MC}(P_{\pi_n!}(a_i), a_i)}_{(|\emptyset|!)(|A|-1)!} \right) \\ &= \frac{1}{|A|!} \left( \underbrace{(\text{MC}(\emptyset, a_i) + \dots + \text{MC}(\emptyset, a_i))}_{(|\emptyset|!)(|A|-1)!} + \dots + \underbrace{(\text{MC}(C, a_i) + \dots + \text{MC}(C, a_i))}_{|C|!(|A|-1-|C|)!} + \dots \right. \\ &\quad \left. \dots + \underbrace{(\text{MC}(A \setminus \{a_i\}, a_i) + \dots + \text{MC}(A \setminus \{a_i\}, a_i))}_{(|A|-1)!(|A|-1-(|A|-1))!} \right) \\ &= \frac{1}{|A|!} \left( (|A|-1)!(\text{MC}(\emptyset, a_i)) + \dots + |C|!(|A|-1-|C|)!(\text{MC}(C, a_i)) + \dots \right. \\ &\quad \left. \dots + (|A|-1)!(\text{MC}(A \setminus \{a_i\}, a_i)) \right) \\ &= \frac{1}{|A|!} \sum_{C \subseteq A \setminus \{a_i\}} |C|!(|A|-|C|-1)! \text{MC}(C, a_i) \end{aligned}$$

Using the Definition 6 we can compute the Shapley value for the game in Example 2.

**Example 3.** We will consider all permutations of the set  $A$  from the game  $g_2$  defined in Example 2. For each permutation we will compute the marginal contribution of the agent  $a_2$  to the set  $P_\pi(a_2)$ .

$a_1, a_2, a_3$	$\text{MC}(\{a_1\}, a_2) = \nu_2(\{a_1, a_2\}) - \nu_2(\{a_1\}) = 200$
$a_1, a_3, a_2$	$\text{MC}(\{a_1, a_3\}, a_2) = 400$
$a_2, a_1, a_3$	$\text{MC}(\emptyset, a_2) = 400$
$a_2, a_3, a_1$	$\text{MC}(\emptyset, a_2) = 400$
$a_3, a_1, a_2$	$\text{MC}(\{a_1, a_3\}, a_2) = 400$
$a_3, a_2, a_1$	$\text{MC}(\{a_3\}, a_2) = 200$

Thus, for the game  $g_2$  we have:

$$\begin{aligned}\phi_2^{SV} &= \frac{200+400+400+400+400+200}{6!} = 333\frac{1}{3} \\ \phi_1^{SV} &= 133\frac{1}{3} \\ \phi_3^{SV} &= 133\frac{1}{3}\end{aligned}$$

Generally, using equation (2.3) computing Shapley value requires iterating through  $2^{|A|}$  subsets of  $A$ , and using equation (2.4) requires iterating through all  $|A|!$  permutations of  $A$ . Furthermore, computing the Shapley value has been shown to be NP-Hard (or even worse, #P-Complete) for many specific classes of games [Deng and Papadimitriou, 1994, Nagamochi et al., 1997].

Therefore, many new succinctly representations were proposed in the literature. The most popular we introduce in Section 2.4 and Section 2.5.

### 2.3. Semivalues

The Shapley value is designed to divide payoff, which is equivalent to evaluating the power of agents in cooperative game. However, if we focus only on the power index of players, we can drop the Efficiency axiom. *Semivalues* introduced by Dubey et al. [1981] represent an important class of power indexes, among which only Shapley value satisfies Efficiency. To define Semivalues, we will use the notion of marginal contribution of the player  $i$  to the coalition  $C$  introduced in Definition 4. Let  $\beta : \{0, 1, \dots, |A| - 1\} \rightarrow [0, 1]$  be a function such that  $\sum_{k=0}^{|N|-1} \beta(k) = 1$ . Intuitively, when we calculate the expected marginal contribution of a node,  $\beta(k)$  will be the probability that a coalition of size  $k$  is chosen for this node to join. This is why  $\beta(k)$  is defined on values ranging from 0 to  $|N| - 1$ . Note that the function  $\beta$  is a discrete

probability distribution. Since a player  $a_i$  cannot join a coalition that it is already in, we only need to look at coalitions not containing  $a_i$ .

**Definition 7 (Semivalue).** Given  $\beta$  and a coalitional game  $(A, \nu)$  the Semivalue for a player  $a_i$  is given by:

$$\phi_i^{SEMI}(A, \nu) = \sum_{0 \leq k < |V|} \beta(k) \mathbb{E}_{C_k}[\mathbf{MC}(C_k, i)], \quad (2.5)$$

where  $C_k$  is the random variable of all possible coalitions of size  $k$  drawn with uniform probability from the set  $A \setminus \{a_i\}$ , and  $\mathbb{E}_{C_k}[\cdot]$  is the expected value operator for the random variable  $C_k$ .

The *Shapley value* (Definition 5) and the *Banzhaf index* of power [Banzhaf, 1965] are two prominent and well-known examples of Semivalues. They are defined by  $\beta$ -functions  $\beta^{Shapley}$  and  $\beta^{Banzhaf}$ , respectively:

$$\beta^{Shapley}(k) = \frac{1}{|A|} \quad \text{and} \quad \beta^{Banzhaf}(k) = \frac{\binom{|A|-1}{k}}{2^{|A|-1}}.$$

All Semivalues satisfy *Symmetry*, *Null player* and *Additivity* axioms and except for the Shapley value, all these solutions violate the *Efficiency* axiom, which makes them of limited use as fair cost sharing values. However, they are widely used as indicators of power in cooperative games [Monderer and Samet, 2002].

## 2.4. Marginal Contribution Networks

The size of the representation of coalitional games in the characteristic function form is exponential in the number of agents and consequently computing Shapley value for such games is generally hard. In order to overcome this limitation the researchers look for more concise representations. One of the most interesting and very intuitive techniques for representing coalitional games was introduced by Jeong and Shoham [2005]. The authors proposed *marginal contribution nets (MC-Nets)*. This representation has a number of desirable properties: it is fully expressive, concise for many characteristic function games of interest, and facilitates a very efficient technique for computing the Shapley value. More specifically, in such representation we can compute the Shapley value in the linear time in respect to the size of a representation.

Let us introduce the syntax of Boolean formulas:

$$\begin{aligned} A &:= a_i \mid \neg a_i & i &\in \{1 \dots n\}, \\ \mathcal{F} &:= (\mathcal{F}) \mid \mathcal{F} \wedge \mathcal{F} \mid \mathcal{F} \vee \mathcal{F} \mid \mathcal{F} \oplus \mathcal{F} \mid A, \end{aligned}$$

where each literal  $a_i$  represents the agent in the game. In other words, this formula is a binary rooted tree with positive or negative literals on leafs and internal nodes labeled with one of the following logical connectives: conjunction ( $\wedge$ ), disjunction ( $\vee$ ) and exclusive disjunction ( $\oplus$ ).

**Definition 8 (General Marginal Contribution Network).** *The coalitional game in the form of General Marginal Contribution Network is a pair  $(A, \mathcal{R})$ , where  $A$  is a set of players, and  $\mathcal{R}$  is the set of rules. Rules are of the form of  $\mathcal{F} \rightarrow V$ , where  $\mathcal{F}$  is a Boolean formula.*

Ieong and Shoham [2005] obtained positive computational results for Marginal Contribution Network, where literals are connected only by conjunction ( $\wedge$ ).

**Definition 9 (Marginal Contribution Network).** *The coalitional game in the form of Marginal Contribution Network is a pair  $(A, \mathcal{R}^*)$ , where  $A$  is a set of players, and  $\mathcal{R}^*$  is the set of rules. Rules are of the form of  $\mathcal{F}^* \rightarrow V$ , where  $\mathcal{F}^*$  is a conjunction of positive or negative literals.*

A coalition  $C$  is said to *meet* a given formula  $\mathcal{F}^*$  if and only if  $\mathcal{F}^*$  evaluates to `true` when the values of all Boolean variables that correspond to players in  $C$  are set to `true`, and the values of all Boolean variables that correspond to players not in  $C$  are set to `false`. We write  $C \models \mathcal{F}^*$  to mean that  $C$  meets  $\mathcal{F}^*$ . More formally:

$$C \models \mathcal{F}^* \iff \forall_{a_i \in \mathcal{F}^*} a_i \in C \text{ and } \forall_{\neg a_i \in \mathcal{F}^*} a_i \notin C$$

In MC-Nets, if coalition  $C$  does not meet any rule then its value is 0. Otherwise, the value of  $C$  is the sum of every  $V$  from the rules of which the  $\mathcal{F}^*$  are met by  $C$ . More formally:

$$v(C) = \sum_{\mathcal{R}^* \ni (\mathcal{F}^* \rightarrow V) \rightarrow V: C \models \mathcal{F}^*} V.$$

Ieong and Shoham showed that, given an MC-Net representation in which rules are made only of conjunctions of positive and/or negative literals, the Shapley value can be computed in time linear in the number of rules. Taking advantage of the additivity property of the Shapley value, every basic rule can be considered separately as a game on its own. Thus, we need to compute the Shapley value for a single rule, and based on this we will be able to compute the Shapley value for the entire game.

**Theorem 1. [Ieong and Shoham, 2005]** *In coalitional game  $(A, \mathcal{R}^*)$  in which  $\mathcal{R}^*$  consists of a single rule  $\mathcal{F}^* \rightarrow V$ , where  $p$  is the number of positive literals and  $n$  is the number of negative ones, the Shapley value for the player  $a_i \in \mathcal{F}^*$  is:*

$$\phi_i^{SV} = \frac{V}{p \binom{p+n}{n}},$$

and for a player  $\neg a_j \in \mathcal{N}$  the Shapley value is:

$$\phi_j^{SV} = \frac{-V}{n \binom{p+n}{p}}.$$

*Proof.* From the Symmetry axiom we know that all players  $a_i \in \mathcal{F}^*$  have the same Shapley value, because they are indistinguishable. This observation also applies to all players  $\neg a_j \in \mathcal{F}^*$ . In this proof we will use Definition 6. We want to count the number of permutations such that for the given player  $a_i \in \mathcal{F}^*$  the  $\text{MC}(P_\pi(a_i), a_i) \neq 0$ . To satisfy this condition all players corresponding to positive literals from  $\mathcal{F}^*$  must appear before  $a_i$  and all players corresponding to negative literals from  $\mathcal{F}^*$  after him. Now, we will construct such permutations:

- Let us choose  $p + n$  positions in the sequence of all elements from  $A$ . We can do this in  $\binom{|A|}{p+n}$  ways.
- Than, in the first  $p$  chosen positions place all players corresponding to positive literals and in the last  $n$  place all players corresponding to negative literals. The number of such line-ups is  $(p-1)!n!$ .
- The remaining players can be arranged in  $(|A| - (p+n))!$  ways.

Thus, we have:

$$\begin{aligned} \phi_i^{SV} &= \frac{1}{|A|!} \sum_{\pi \in \Pi(A)} \text{MC}(P_\pi(a_i), a_i) \\ &= \frac{1}{|A|!} \binom{|A|}{p+n} (p-1)!n! (|A| - (p+n))! V \\ &= \frac{(p-1)!n!}{(p+n)!} V \\ &= \frac{V}{p \binom{p+n}{n}} \end{aligned}$$

The proof for  $a_j \in \mathcal{N}$  is analogous. ■

Now, let us consider the game  $g_2$  from Example 2 and transform its representation from characteristic function form to Marginal Contribution Nets. Next, we will compute the Shapley value based on the above theorem.

**Example 4.** Firstly, we define the set of rules  $\mathcal{R}^*$  that corresponds to the characteristic function  $v_2$  from the game  $g_2$ .

$$\begin{aligned} r_1 : & \quad \quad \quad a_2 \rightarrow 400 \\ r_2 : & \quad \quad a_1 \wedge \neg a_2 \rightarrow 200 \\ r_3 : & \quad a_3 \wedge \neg a_1 \wedge \neg a_2 \rightarrow 200 \\ r_4 : & \quad \quad a_1 \wedge a_2 \wedge a_3 \rightarrow 200 \end{aligned}$$

Secondly, in order to compute  $v_2(\{a_1, a_2, a_3\})$  we need to iterate through all rules that this coalition meets. We see that  $\{a_1, a_2, a_3\} \models a_2$  and  $\{a_1, a_2, a_3\} \models a_1 \wedge a_2 \wedge a_3$ , so



$\nu_2(\{a_1, a_2, a_3\}) = 400 + 200 = 600$ . In the same manner we can compute the value of any other coalition. Now, let us consider the Shapley value vector for each game defined by a single rule.

$$\begin{aligned} r_1 : \quad & (0, \frac{400}{1\binom{1}{1}}, 0) = (0, 400, 0) \\ r_2 : \quad & (\frac{200}{1\binom{2}{1}}, \frac{-200}{1\binom{2}{1}}, 0) = (100, -100, 0) \\ r_3 : \quad & (\frac{-200}{2\binom{3}{1}}, \frac{-200}{2\binom{3}{1}}, \frac{200}{1\binom{3}{2}}) = (-33\frac{1}{3}, -33\frac{1}{3}, 66\frac{2}{3}) \\ r_4 : \quad & (\frac{200}{3\binom{3}{0}}, \frac{200}{3\binom{3}{0}}, \frac{200}{3\binom{3}{0}}) = (66\frac{2}{3}, 66\frac{2}{3}, 66\frac{2}{3}) \end{aligned}$$

We sum up all vectors, and we obtain the final solution of this game:

$$\begin{aligned} \phi_1^{SV} &= 0 + 100 - 33\frac{1}{3} + 66\frac{2}{3} = 133\frac{1}{3} \\ \phi_2^{SV} &= 400 - 100 - 33\frac{1}{3} + 66\frac{2}{3} = 333\frac{1}{3} \\ \phi_3^{SV} &= 0 + 0 + 66\frac{2}{3} + 66\frac{2}{3} = 133\frac{1}{3} \end{aligned}$$

To conclude the above example, we have obtain the same results as in Example 2 but in order to compute the Shapley value for each player we have performed less operation than when iterating through all permutations. However, creating rules only with conjunctions is not an effective way for many games. Thus, in the next section we will introduce more effective representation of the coalitional games.

## 2.5. Read-once Marginal Contribution Networks

Many coalitional games can be represented more succinct than with Marginal Contribution Networks (Definition 8). Elkind et al. 2009a defined a class of rules, called *read-once MC-Net rules*, which are considerably more succinct than basic rules, while enjoying similarly attractive computational properties. The syntax of read-once Boolean formulas is the same as in General Marginal Contribution Networks, but each literal  $a_i$  can only appear once. In other words, this formula is also a binary rooted tree with positive or negative literals on leafs and internal nodes labeled with one of the following logical connectives: conjunction ( $\wedge$ ), disjunction ( $\vee$ ) and exclusive disjunction ( $\oplus$ ).

**Definition 10 (Read-once Marginal Contribution Network).** *The coalitional game in the form of Read-once Marginal Contribution Network is a pair  $(A, \mathcal{R})$ , where  $A$  is a set of players, and  $\mathcal{R}$  is the set of rules. Rules are of the form of  $\mathcal{F} \rightarrow V$ , where  $\mathcal{F}$  is read-once Boolean formulas.*

Elkind et al. [2009a] proved that, in General Marginal Contribution Networks, where each literal can appear in a formula many times, the problem of computing the Shapley value is

intractable (i.e., NP-hard). However, for the read-once Boolean formulas the authors presented a polynomial-time algorithm to compute the Shapley value.

In Chapter 8 we extend the positive results obtained by Elkind et al. [2009a] to generalized coalitional games.

## 2.6. Graph Theory

The notion of a *graph* (or a *network*) will play a the key role in this dissertation. These mathematical entities models many real-life problems from different disciplines, such as biology, chemistry, or sociology [Barabási, 2003].

**Definition 11 (Graph).** *The simple graph  $G$  is a pair  $(V, E)$ , where  $V = \{v_1, v_2, \dots, v_{|V|}\}$  is the set of nodes (or vertices), and  $E = \{e_1, e_2, \dots, e_{|E|}\}$  is the set of edges (or links). Edges  $e = (v_i, v_j)$  are a two-elements sets of nodes from  $V$ .*

In the same manner, but with different definition of edge, we define the directed graph.

**Definition 12 (Directed graph).** *The directed graph  $G$  is a pair  $(V, E)$ , where  $V = \{v_1, v_2, \dots, v_{|V|}\}$  is the set of node, and  $E = \{e_1, e_2, \dots, e_{|E|}\}$  is the set of edges. Edges  $e = (v_i, v_j)$  are a two-elements ordered pairs of nodes from  $V$ .*

In such graphs we say that an edge  $(v_i, v_j)$  is directed from the node  $v_i$  to the node  $v_j$ .

**Definition 13 (Weighted graph).** *The weighted graph  $G = (V, E, \lambda)$  is a simple/directed graph, with a function  $\lambda(v, u) : E \rightarrow \mathbb{R}$  that evaluates each edge. The value  $\lambda(e)$  is called the wight of an edge  $e$ .*

A *path*  $p_{st}$  from node  $s$  to node  $t$  in a graph  $G$  is an ordered set  $(v_0, v_1, \dots, v_n)$  such that  $v_0 = s$  and  $v_n = t$  and  $(v_i, v_{i+1}) \in E$  for all  $i$  with  $1 \leq i < n$ . We define the set of *neighbours* of a node  $v$  and a subset  $C$  of nodes by  $N(v_i) = \{v_j : (v_i, v_j) \in E\}$  and  $N(C) = \bigcup_{v_i \in C} N(v_i) \setminus C$ , respectively. We refer to the *degree* of a node  $v$  by  $\deg(v) = |N(v)|$ , and the *degree* of a set  $C$  as  $\deg(C) = |N(C)|$ . The *distance* from a node  $s$  to a node  $t$  is denoted by  $\text{dist}(s, t)$  and is defined as the size of the shortest path between  $s$  and  $t$ . The distance between a node  $v$  and a subset of nodes  $C \subseteq V$  is denoted by  $\text{dist}(C, v) = \min_{u \in C} \text{dist}(u, v)$ .

For a further comprehensive introduction to graph theory the reader is kindly referred to the work by Cormen 2001.

## 2.7. Centrality Measures

One of the important aspects of research on networks is to determine which nodes and edges are more critical (or *central*) to the functioning of the entire network than others. For example,

one may want to identify the focal hubs in a road network [Schultes and Sanders, 2007], the most critical functional entities in a protein network [Jeong et al., 2001], or the most influential people in a social network [Kempe et al., 2003]. To this end, various centrality metrics, such as *degree*, *closeness*, *eigenvalue* or *betweenness*, have been extensively studied in the literature.<sup>1</sup>

**Definition 14 (Centrality).** For a given graph  $G = (V, E)$  the classic (also standard) centrality is a function  $c : V \rightarrow \mathbb{R}$  that for each node  $v \in V$  it determines the importance of this node measured in real value.

Generally speaking, centrality analysis aims to create a consistent ranking of nodes within a network. To this end, *centrality measures* assign a score to each node that in some way corresponds to the importance of that node given a particular application. Since “importance” depends on the context of the problem at hand, many different centrality measures have been developed. The three well-known and widely applied measures introduced by Freeman [1979] are: *degree centrality*, *closeness centrality* and *betweenness centrality*.

**Definition 15 (Degree centrality).** For a given graph  $G = (V, E)$  degree centrality is a function  $c_d : V \rightarrow \mathbb{R}$ :

$$c_d(v) = \deg(v)$$

where  $\deg(v)$  is a degree of the node  $v$ .

The intuition standing behind this measure is that it reflects the activity of a given node. This activity can be expressed as a number of friends in social portal, or number of articles published together with other peers in science collaboration network.

**Definition 16 (Betweenness centrality).** For a given graph  $G = (V, E)$  betweenness centrality is a function  $c_b : V \rightarrow \mathbb{R}$ :

$$c_b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where  $\sigma_{st}$  is the number of shortest paths between  $s$  and  $t$  (if  $s = t$  then  $\sigma_{st} = 1$ ), and  $\sigma_{st}(v)$  is the number of shortest paths between  $s$  and  $t$  passing through a node  $v$  (if  $v \in \{s, t\}$  then  $\sigma_{st}(v) = 0$ ).<sup>2</sup>

This measure and its different variants is generally used to indicate the most important nodes in the process of controlling information flow. For instance it can be used to determine the most important routers in computer network.

**Definition 17 (Closeness centrality).** For a given graph  $G = (V, E)$  closeness centrality is a function  $c_c : V \rightarrow \mathbb{R}$ :

$$c_c(v) = \sum_{s \neq v} d(s, v)$$

<sup>1</sup> Koschützki et al. [2005] presented an overview of the most important centrality metrics.

<sup>2</sup> To deal with unconnected graphs it is assumed that  $\frac{0}{0} = 0$ .

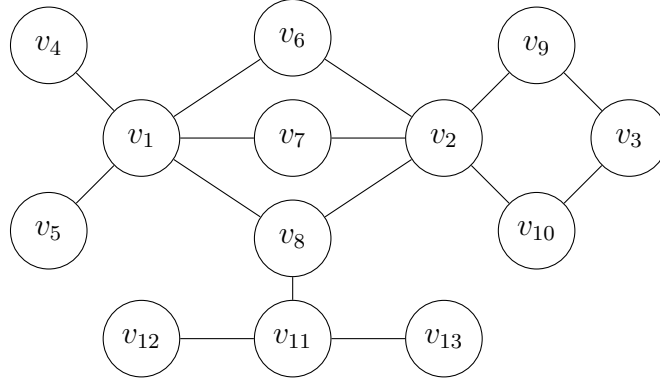


Figure 2.1: Sample network of 13 nodes

where  $\text{dist}(s, v)$  is a distance between nodes  $s$  and  $v$ .

This centrality has two aspects: one that evaluates the most important nodes in the process of discrimination of information, and one that evaluates the most independent nodes. For instance, if we want to construct an efficient network of communication in our company, the CEO should be in the middle of this structure that his orders would immediately hit the targets. On the other hand, the person with the lowest closeness centrality is the most independent person, because the information coming to him from others requires the smallest number of brokers. For unconnected networks this measure is sometimes presented as a sum of fractions:  $\sum_{s \neq v} \frac{1}{d(s, v)}$ .

In the sample network in Figure 2.1, nodes  $v_1$  and  $v_2$  have degree 5 and, if judged by degree centrality, these are the most important nodes within the entire network ( $c_d(v_1) = c_d(v_2) = 5$ ). Conversely, closeness centrality focuses on distances among nodes and gives high value to the nodes that are close to all other nodes. With this measure, node  $v_8$  in Figure 2.1 is ranked top with lowest value  $c_c(v_8) = 22$ . The last measure—betweenness centrality—considers shortest paths (i.e., paths that use the minimal number of links) between any two nodes in the network. The more shortest paths the node belongs to, the more important it is. With this measure,  $c_b(v_2) = 32$  and the node  $v_2$  in Figure 2.1 is more important than all the other nodes (including  $v_1$  and  $v_8$ , which are chosen by other measures as the most important node). Clearly, all these measures expose different characteristics of a node. Consider, for instance, an epidemiology application, where the aim is to identify those people (i.e., nodes) in the social network who have the biggest influence on the spread of the disease and should become a focal point of any prevention or emergency measures. Here, degree centrality ranks top nodes with the biggest immediate *sphere of influence*—their infection would lead to the highest number of adjacent nodes being exposed to the disease. On the other hand, closeness centrality identifies those nodes whose infection would lead to the *fastest* spread of the disease throughout the society. Finally, betweenness centrality reveals the nodes that play a crucial role in *passing* the disease from one person in a network to another.

## 2.8. Community Structures

Apart from centrality measures, the other important aspect of networks extensively studied in literature is their community structure. Intuitively, a *community* (also called a cluster) is a group of nodes in a network such that there are comparatively many edges among the nodes within the group and fewer with nodes outside the group. The communities can be non-overlapping, which mean that node can belong to only one community, and overlapping. In this dissertation we focus on non-overlapping communities. The graph divided into communities forms a *community structure*.

**Definition 18 (Community structure).** For a given graph  $G = (V, E)$  the community structure  $CS = \{C_1, C_2, \dots, C_m\}$  is the partition of the set  $V$ , which implies that  $\emptyset \notin CS$ ,  $\sum_{C \in CS} C = V$  and  $\forall_{C_1, C_2 \in CS} C_1 \neq C_2 \implies C_1 \cap C_2 = \emptyset$ .

Community detection is a key research topic in network analysis [Fortunato, 2010]. It is relevant to many fields, where graph representations are common, including sociology [Girvan and Newman, 2002], biology [van Laarhoven and Marchiori, 2012] and computer science [Misra et al., 2012]. Community detection raises various conceptual question. One such question is how to evaluate a given division of a network into communities. One of the prominent method of doing this was introduced by Newman [2006] and is called *modularity*.

**Definition 19 (Modularity).** For a given graph  $G = (V, E)$  and community structure  $CS = \{C_1, C_2, \dots, C_m\}$  the modularity is:

$$Q(G, CS) = \sum_{C \in CS} \left( \frac{|E(C)|}{|E|} - \left( \frac{\#deg(C)}{2|E|} \right)^2 \right), \quad (2.6)$$

where  $E(C) = \{(u, v) \in E : u, v \in C\}$ , and  $\#deg(C) = \sum_{v \in C} deg(v)$ .

The first term of the above equation is the fraction of edges inside community  $C$  and the second term, in contrast, is the expected fraction of edges in that community under the assumption that edges were located at random in the network, in which the degrees of nodes remains unchanged in respect to the original graph.

## 3. Related Work

In this dissertation we are concerned with computational aspects of game-theoretic centrality measures. Hence, there are three main bodies of related literature: on centrality measures (Section 3.1), on game-theoretic centrality measures (Section 3.2), and on computational aspects of coalitional games (Section 3.3). The related work on the more specific subjects will be presented in the chapters corresponding to these topics.

### 3.1. Centrality Measures and Their Classification

Centrality measures constitute one of the main research directions in the literature on network analysis. In his classic work, Freeman [1979] proposed and formalized three different concepts of centrality: (i) *degree centrality* is based on being directly connected to other nodes; (ii) *closeness centrality* is based on being close to all other nodes; and (iii) *betweenness centrality* is based on lying between other nodes. Importantly, these three centralities are build upon three conceptually different properties extracted from networks: connections, distances and geodesics, respectively, and all three can be easily used as a basic ground for various extensions. To name only the few, Newman [2004] proposed weighted degree centrality, a appropriate measure of degree for weighted networks, Brandes [2008] introduced distance-scaled betweenness centrality, and Boldi and Vigna [2013] harmonic centrality, the extension of closeness one.

The fourth classical measure is *eigenvector centrality*, which is based on the idea that connections to more important nodes should be valued more than otherwise equal connections to less important nodes [Bonacich, 1972]. One of the most known extension of this measure was proposed by Brin and Page [1998], who proposed famous *PageRank* metric

The topics closely related to centralities are that of the *percolation* on complex networks and *social capital*. Regarding percolation, Callaway et al. [2000] studied different models of nodes failures, where each element of the graph have some probability of being disabled. Such studies gives answer to the problem how a given network is vulnerable to a random node or edge failure. Here the centralities can be used to determine the most important nodes to protect. Such studies were done by Holme et al. [2002], who compared degree and betweenness centralities. He postulated that betweenness centrality is a good basis for strategies of protecting nodes against random nodes failures.

The "social capital" term is much wider than "centrality" and is used to describe both "*the value of an individual's social relationships*" [Burt, 1992] and "*a quality of groups*" [Everett and Borgatti, 1999] within the society. Borgatti et al. [1998] propose to classify social capital along two dimensions. The first is whether it concerns individual actor or a group, or whether we focus on internal or external links (see Table 3.1).

Type of Node	Type of Focus	
	internal	external
Individual	(A) —	(B) standard centralities [Freeman, 1979]
Group	(C) centralisation [Freeman, 1979]	(D) group centralities [Everett and Borgatti, 1999]

Table 3.1: Classification by Borgatti et al. [1998] of different conceptions (forms) of social capital (including different *centralities*), with examples of such measures.

This classification distinguishes three fields of social capital studies (or centrality): **(A)** in this category are examined the internal properties of autonomous actors and this is not part of the studies on social capital nor social networks analysis; **(B)** in this field the individual actor is examined in the context of its relations to external world and most standard centralities belong to this category; **(C)** in this category the group of actors are studied in the context of internal relations among them; **(D)** the last type of social capital is engaged in studying group of actors and their relationships with external world. For all types of social capital many different measures have been proposed (see Table 3.1) and most of them the literature refers to as *centralities*. Although the three categories of social capital are considered separately, there are cases where they influence each other. For instance, in situations where actors are divided into groups, the role of each individual strictly depends on the role of the entire group and the role of the group simply depends on its members.

Social capital is related to social networks describing people and their relationships. Centrality measures can be used to analyze any kind of networks.

### 3.2. Game-theoretic Approach to Centrality Measures

Game-theoretic centrality is an extension of standard centrality based on cooperative game theory. The basic idea behind game-theoretic centrality is to consider the nodes of the network to be a *players* in a cooperative game, where the payoffs of coalitions are derived from the network topology. Such an analogy between graph and game theory allows to use various game-theoretic solution concepts as a centrality measure of individual nodes.<sup>1</sup>

<sup>1</sup> See the next chapter for more details.

The first author to introduce game-theoretic centrality were Grofman and Owen [1982], who used *the Banzhaf index of power* [Banzhaf, 1965]. This is a widely-used solution concept that quantifies the power of a player by counting the number of coalitions in which this player is a *swing player*—a player whose addition to the coalition changes it from being unsuccessful (e.g., in achieving some goal) to being successful. Grofman and Owen mapped a *simple coalitional game*, a coalitional game where every coalition has a value of 1 or 0, onto a directed graph by considering paths in the graph to be coalitions. Here, a swing player is a node that is crucial to maintain communication among the nodes in the path. Then, Grofman and Owen’s Banzhaf index-based centrality says that the more times a node is a swing node within all possible paths of nodes, the more important this node is in the entire network.

The game-theoretic centrality were reinvented by Gómez et al. [2003], who proposed a slightly different approach. Instead of restricting themselves to simple coalitional games, the authors study arbitrary games where a coalition’s value can be any real number. In this context, the authors propose a centrality measure based on the two prominent solution concepts: *the Shapley value* [Shapley, 1953] and *the Myerson value* [Myerson, 1977].

The first value is arguably the best normative solution concept known to date, and was discussed in Section 2.2. *The Myerson value* is based on *graph-restricted games*, where the underlying assumption is that only connected coalitions (i.e., the group of nodes that between any pair exists a path passing through only members of this group) can communicate and create any value added. Myerson assumed that the value of disconnected coalition would simply be the sum of the values of its connected components. A celebrated result of Myerson for this setting is a dedicated solution concept—closely related to the Shapley value—called the Myerson value. In fact, it is the Shapley value, but evaluated for the characteristic function sensible to coalitions connectivity, as was described above.

Taking Shapley value and Myerson value, Gómez et al. [2003] defined the centrality of a node  $v_i$  to be the difference between the Shapley value of  $v_i$  (which is independent of the network connectedness) and the Myerson value of  $v_i$  (which takes the network connectedness into consideration). Thus, by interpreting the Shapley value as a power index in a coalitional game, Gómez et al.’s Shapley/Myerson-based centrality measure represents the increase (or decrease) in  $v_i$ ’s power due to its position in the network.

This direction of research took also del Pozo et al. [2011], who proposed a measure for directed graphs, based on *generalised coalitional games* [Nowak and Radzik, 1994], where the value of a coalition is influenced by the sequence in which the players have joined it. More specifically, the authors adapted these games to networks in the spirit of Myerson and then based their measure on a parametric family of solution concepts that include two extensions of the Shapley value to generalised coalitional games: one by Nowak and Radzik [1994] and



the other by Sánchez and Bergantiños [1999]. In a similar spirit, Amer et al. [2007] defined a game-theoretic centrality measure called accessibility in oriented networks.

A computational study of game-theoretic centrality measures based on connectivity games on networks (inspired by Myerson) was done in Chapter 7. We showed that the efficient computation of the Shapley value is impossible even in the case of a very simple definition of the connectivity game on graph. We also proposed algorithm that significantly improve the computation of such game-theoretic centrality measures. We need to mention that our algorithm was further improved by Skibski et al. [2014] and modified and applied to analyze social networks by Narayanam et al. [2014].

All the above works are inspired in one way or the other by the graph-restricted games in the spirit of Myerson [1977], where the value of a group of nodes is mainly determined by connectedness of this group.

This dissertation contributes to a parallel stream of literature in which coalitions of nodes are not graph-restricted, i.e., their values do not depend directly on whether they are connected or not. The origins of this approach can be traced back to the concept of *group centrality* introduced by Everett and Borgatti [1999] which takes into account synergies among nodes. By computing group centrality for every possible group of nodes (not necessarily connected) we obtain a well-defined coalitional game. Next, one can apply a solution concept, such as the Shapley value or the Banzhaf index, to evaluate each individual nodes. Such evaluation attribute synergies to individual nodes, that were captured by group centrality measure.

The measure that captured the ability of nodes to dominate other nodes in directed networks was proposed by van den Brink and Borm [2002]. The authors do not refer to their measure as *centrality*, but from the social network analysis perspective it is. van den Brink and Gilles [2000] continue this line of research by analysing these measures more deeply and, more importantly, by providing an axiomatic approach to them. These are the first works to use general Shapley value approach to game-theoretic centralities.

This second line of research was also used by Suri and Narahari [2010] in the interesting application of influence propagation in networks. Specifically, the authors constructed a coalitional game by assigning to coalitions payoffs equal to their group degree. Then, the authors used the Shapley value to approximate the solution of the top  $k$ -node problem, i.e., the problem of identifying the  $k$  most influential nodes in the network. This approach was followed up by Adamczewski et al. [2014], who studied different versions of group degree.

### 3.3. Computational Aspects of Coalitional Games – Representations

Let us now turn our attention to the computational issues. In more detail, the computational challenges related to the game-theoretic solution concepts that are tackled in this dissertation

have been extensively studied in the literature on algorithmic aspects of coalitional games. Since the path-breaking work by Deng and Papadimitriou [1994], a considerable attention has been paid to developing more efficient *representations* of games. Such representation can be divided into two main categories [Wooldridge and Dunne, 2006].

The first category, to which also this dissertation indirectly contributes, models the characteristic function using specific combinatorial structures such as graphs. This line of research include works by Deng and Papadimitriou [1994], Greco et al. [2009], Wooldridge and Dunne [2006], and the representations discussed by Aziz and de Keijzer [2011]. Such representations are guaranteed to be succinct, however they can express only certain games. In our case, the characteristic function is the function of the group betweenness centrality.

In the second category of representations the emphasis is placed on full expressiveness, often at the expense of succinctness. These representations include MC-Nets [Jeong and Shoham, 2005], its read-once extension [Elkind et al., 2009b], Synergy Coalition Groups [Conitzer and Sandholm, 2006], the Decision-Diagrams-based representations [Aadithya et al., 2011, Sakurai et al., 2011], the vector-based representation [Tran-Thanh et al., 2013], and the representation for graph-restricted weighted voting games [Skibski et al., 2015] .

We note that even for some succinctly representable games, computing the Shapley value and other game-theoretic solution concepts can be challenging (NP-Hard or even #P-Complete). This is the case with weighted voting games [Deng and Papadimitriou, 1994], threshold network flow games [Bachrach and Rosenschein, 2009] and minimum spanning tree games [Nagamochi et al., 1997]. Also Aziz et al. [2009] proved negative results for the Shapley-Shubik power index for the spanning connectivity games on multigraphs, and Bachrach et al. [2008] for the Banzhaf index for certain class of connectivity games.

Perhaps the most widely known positive results are those aforementioned due to Deng and Papadimitriou [1994] and Jeong and Shoham [2005]. In particular, Deng and Papadimitriou proposed a representation based on graphs, weighted edges of which represent the marginal contribution of both adjacent nodes (agents) to any coalition they belong to. While this representation is clearly not fully expressive, the Shapley value can be computed in time linear in the size of the graph. The representation of Jeong and Shoham [2005] consists of a finite set of logical rules of the following form: *Boolean Expression*  $\rightarrow$  *Real Number*, where agents are represented by atomic boolean variables (see Definition 9). The value of a coalition is equal to the sum of the Real Numbers in those rules whose that are satisfied by the coalition. Unlike the representation by Deng and Papadimitriou, the representation by Jeong and Shoham is fully expressive and, for certain types of Boolean Expressions, it allows for the Shapley value to be computed in time linear in the size of the representation.

The number of authors proposed alternative representations of coalitional games: coalitional skill games [Bachrach and Rosenschein, 2008] or synergy coalitional groups [Aadithya et al.,

2011], which allow for representing input compactly for certain games and admit algorithms (e.g. for computing the Shapley value) that run in polynomial time in the size of such a compact input. Finally, one can find in the literature the representation of the extension of coalitional games to games with externalities [Michalak et al., 2009, 2010].

## 4. Polynomial Algorithms for Game-theoretic Centrality Measures

The Shapley value has recently been advocated as a useful measure of centrality in networks. However, although this approach has a variety of real-world applications (including social and organisational networks, biological networks and communication networks), its computational properties have not been widely studied. To date, the only practicable approach to compute Shapley value-based centrality has been via Monte Carlo simulations which are computationally expensive and not guaranteed to give an exact answer. Against this background, in this chapter we present the study of the computational aspects of the Shapley value for network centralities. Specifically, we develop exact analytical formulae for Shapley value-based centrality in both weighted and unweighted networks and develop efficient (polynomial time) and exact algorithms based on them. We empirically evaluate these algorithms on two real-life examples (an infrastructure network representing the topology of the Western States Power Grid and a collaboration network from the field of astrophysics) and demonstrate that they deliver significant speedups over the Monte Carlo approach. For instance, in the case of unweighted networks our algorithms are able to return the exact solution about 1600 times faster than the Monte Carlo approximation, even if we allow for a generous 10% error margin for the latter method.

Against this background, in this chapter we present the study of the computational aspects of the Shapley value for network centralities. Specifically, we develop exact analytical formulae for Shapley value-based centrality in both weighted and unweighted networks and develop efficient (polynomial time) and exact algorithms based on them. We empirically evaluate these algorithms on two real-life examples (an infrastructure network representing the topology of the Western States Power Grid and a collaboration network from the field of astrophysics) and demonstrate that they deliver significant speedups over the Monte Carlo approach. For instance, in the case of unweighted networks our algorithms are able to return the exact solution about 1600 times faster than the Monte Carlo approximation, even if we allow for a generous 10% error margin for the latter method.

The contribution of the author of this dissertation mostly covers the experimental part of this article. More specifically, Algorithms 1-5 were developed by Karthik Aaditha, and formalized

together with the author of this thesis. The Propositions 1 and 2 were proved by the author of this thesis. Also Algorithms 6-10 were developed by the author of this thesis.

## 4.1. Why Use Game-Theoretic Centrality Measures?

In many network applications, it is important to choose which nodes and edges are the most important. To this end, the concept of *centrality* (see Section 2.7), which aims to quantify the importance of individual nodes/edges, has been extensively studied in the literature [Koschützki et al., 2005, Brandes and Erlebach, 2005].

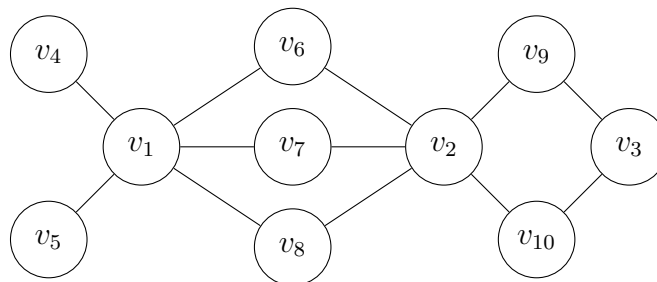


Figure 4.1: Sample network of 10 nodes

Generally speaking, centrality analysis aims to create a *consistent ranking* of nodes within a network. To this end, *centrality measures* assign a score to each node that in some way corresponds to the importance of that node given a particular application. Since “importance” depends on the context of the problem at hand, many different centrality measures have been developed. Three of the most well-known that were introduced in Section 2.7: *degree centrality*, *closeness centrality* and *betweenness centrality*. Recall that we refer to these measures as *classic/standard centralities* (Definition 14).

The common feature of all standard measures is that they assess the importance of a node by focusing only on the role that a node plays by itself. However, in many applications such an approach is inadequate because of *synergies* that may occur if the functioning of nodes is considered in groups. Referring to Figure 4.1 and epidemiology example, a vaccination of individual node  $v_6$  (or  $v_7$  or  $v_8$ ) would not prevent the spread of the disease from the left to the right part of the network (or vice versa). However, the simultaneous vaccination of  $v_6$ ,  $v_7$  and  $v_8$  would achieve this goal. Thus, in this particular context, nodes  $v_6$ ,  $v_7$  and  $v_8$  do not play any significant role individually, but together they do. To quantify the importance of such groups of nodes, the notion of *group centrality* was introduced by Everett and Borgatti [1999]. Intuitively, group centrality works broadly the same way as standard centrality, but now the focus is on the functioning of a given group of nodes, rather than individual nodes. For instance, in Figure 4.1, the group degree centrality of  $\{v_1, v_2\}$  is 7 as they both have 7 distinct adjacent nodes.

Although the concept of group centrality addresses the issue of synergy between the functions is played by various nodes, it suffers from a fundamental deficiency. It focuses on particular, a priori determined, groups of nodes and it is not clear how to construct a consistent ranking of *individual* nodes using such group results. Specifically, should the nodes from the most valuable group be ranked top? Or should the most important nodes be those which belong to the group with the highest average value per node? Or should we focus on the nodes which contribute most to every coalition they join? In fact, there are very many possibilities to choose from.

A framework that does address this issue is the *game theoretic network centrality measure*. In more detail, it allows the consistent ranking of individual nodes to be computed in a way that accounts for various possible synergies occurring within possible groups of nodes [Grofman and Owen, 1982, Gómez et al., 2003]. Specifically, the concept builds upon cooperative game theory—a part of game theory in which agents (or players) are allowed to form coalitions in order to increase their payoffs in the game. Now, one of the fundamental questions in cooperative game theory is how to distribute the surplus achieved by cooperation among the agents. To this end, Shapley proposed to remunerate agents with payoffs that correspond to their individual marginal contributions to the game (Definition 5). In more detail, for a given agent, such an individual marginal contribution is measured as the weighted average marginal increase in the payoff of any coalition that this agent could potentially join. Shapley famously proved that his concept—known since then as the Shapley value—is the only division scheme that meets certain desirable normative properties. Given this, the key idea of the game theoretic network centrality is *to define a cooperative game over a network* in which agents are the nodes, coalitions are the groups of nodes, and payoffs of coalitions are defined so as to meet requirements of a given application. This means that the Shapley value of each agent in such a game can then be interpreted as a *centrality measure* because it represents *the average marginal contribution made by each node to every coalition of the other nodes*.<sup>1</sup> In other words, the Shapley value answers the question of how to construct a consistent ranking of individual nodes once groups of nodes have been evaluated.

In more detail, the Shapley value-based approach to centrality is, on one hand, much more *sophisticated* than the conventional measures, as it accounts for any group of nodes from which the Shapley value derives a consistent ranking of individual nodes. On the other hand, it confers a high degree of *flexibility* as the cooperative game over a network can be defined in a variety of ways. This means that *many different versions of Shapley value-based centrality* can be developed depending on the particular application under consideration, as well as on the features of the network to be analyzed. As a prominent example, in which a specific Shapley value-based

---

<sup>1</sup> We note that other division schemes or power indices from cooperative game theory can also serve as a good solution.

centrality measure is developed that is crafted to a particular application, consider the work of Suri and Narahari [2010] who study the problem of selecting the top- $k$  nodes in a social network. This problem is relevant in all those applications where the key issue is to choose a group of nodes that together have the biggest *influence* on the entire network. These include, for example, the analysis of co-authorship networks, the diffusion of information, and viral marketing. As a new approach to this problem, Suri and Narahari define a cooperative game in which the value of any group of nodes is equal to the number of nodes within, and adjacent to, the group. In other words, it is assumed that the agents' *sphere of influence* reaches the immediate neighbors of the group. Whereas the definition of the game is a natural extension of the (group) degree centrality discussed above, the *Shapley value of nodes in this game* constitutes a new centrality metric that is, arguably, qualitatively better than standard degree centrality as far as the node's influence is concerned. The intuition behind it is visible even in our small network in Figure 4.1. In terms of influence, node  $v_1$  is more important than  $v_2$ , because it is the only node that is connected to  $v_4$  and  $v_5$ . Without  $v_1$  it is impossible to influence  $v_3$  and  $v_4$ , while each neighbor of  $v_2$  is accessible from some other node. Thus, unlike standard degree centrality, which evaluates  $v_1$  and  $v_2$  equally, the centrality based on the Shapley value of the game defined by Suri and Narahari recognizes this difference in influence and assigns a higher value to  $v_1$  than to  $v_2$ .

Unfortunately, despite the advantages of Shapley value-based centrality over conventional approaches, efficient algorithms to compute it have not yet been developed. Indeed, given a network  $G(V, E)$ , where  $V$  is the set of nodes and  $E$  the set of edges, using the original Shapley value formula involves computing the marginal contribution of every node to every coalition which is  $O(2^{|V|})$ . Such an exponential computation is clearly prohibitive for bigger networks (of, e.g, 100 or 1000 nodes). For such networks, the only feasible approach currently outlined in the literature is Monte-Carlo sampling [Suri and Narahari, 2010, Castro et al., 2009]. However, this method is not only inexact, but can be also very time-consuming (see Section 4.4).

The computational challenges are hard to overcome, but fortunately in this dissertation we develop the number of algorithms allowing to compute game-theoretic centralities in polynomial time.

## 4.2. General Definition of Game-Theoretic Network Centrality

In this section we provide the formal definitions of group centralities and game-theoretic network centralities.

In order to formalize the game-theoretic network centrality firstly we need to define the group centralities. These measured introduced by Everett and Borgatti [1999] are the natural extension of standard centralities (Section 2.7) to evaluate group of nodes. Everett and Borgatti introduced three requirements that group centrality should meet. Firstly, these metrics should

be build upon existing concepts of centralities. Secondly, the group centrality applied to the group containing a single individual should yield the same result as standard centrality of this individual. Finally, the group centrality must stem from the relationships between individuals, not between groups. In other words, the value of the group  $C \subseteq V$  of nodes is an intrinsic property of this group and its relationships, and the other group of nodes has no impact on it as long as the relationships of  $C$  are not changed.

**Definition 20 (Group centrality).** *For a given graph  $G = (V, E)$  the group centrality is a function  $c_g : 2^V \rightarrow \mathbb{R}$  that for each subset of nodes  $C \subseteq V$  determines the importance of this group measured in real value.*

Now, the extension of degree, centrality (Definition 15) is:

**Definition 21 (Group degree centrality).** *For a given graph  $G = (V, E)$  and a group of nodes  $C \subseteq V$  group degree centrality is a function  $c_{gd} : 2^V \rightarrow \mathbb{R}$ :*

$$c_{gd}(C) = \deg(C)$$

where  $\deg(C)$  is a degree of the set  $C$  (see Section 2.6).

The extension of betweenness centrality (Definition 16) is:

**Definition 22 (Group betweenness centrality).** *The group betweenness centrality of the set of nodes  $C \subseteq V$  is defined as a function  $c_{gb} : 2^V \rightarrow \mathbb{R}$  such that:*

$$c_{gb}(C) = \sum_{\substack{s \notin C \\ t \notin C}} \frac{\sigma_{st}(C)}{\sigma_{st}},$$

where  $\sigma_{st}(C)$  is the number of shortest paths from  $s$  to  $t$  passing through at least one vertex in  $C$  (if  $s \in C$  or  $t \in C$  then  $\sigma_{st}(S) = 0$ ), and  $\sigma_{st}$  is the number of all shortest paths between  $s$  and  $t$ .

Eventually, the extension of closeness centrality (Definition 17) is defined as follows:

**Definition 23 (Group closeness centrality).** *For a given graph  $G = (V, E)$  and a group of nodes  $C \subseteq V$  closeness centrality is a function  $c_{gc} : 2^V \rightarrow \mathbb{R}$ :*

$$c_{gc}(C) = \sum_{s \notin C} d(C, s)$$

where  $\text{dist}(C, s)$  is a distance between group  $C$  and node  $s$  (see Section 2.6).



In this dissertation, we will be looking at cooperative games on graphs, with the set of players  $V$  for some  $(V, E) = G \in \mathcal{G}$ , where  $\mathcal{G}$  is an infinite set of all graphs.<sup>2</sup> A coalition of players  $C$  will simply be any subset of  $C \subseteq V$ . The characteristic function,  $\nu_G \in \mathcal{V}_G$ , will be any function  $\nu_G : 2^{V(G)} \rightarrow \mathbb{R}$ , but in this work we focus on the group centralities, so we have  $\nu_G = c_g$ . For the purposes of defining a cooperative game on any network, we will systematically associate characteristic functions with graphs through *representation functions*.

**Definition 24 (Representation function).** A representation function is a function  $\psi : \mathcal{G} \rightarrow \mathcal{V}_G$  that maps every graph  $G = (V, E)$  onto a cooperative game  $(N, \nu_G) \in \mathcal{V}_G$  with  $N = V$ .

Now, solution concepts like Shapley value can be used in the network setting by applying them to the characteristic function that a network represents.

**Definition 25 (Game-theoretic centrality measure).** Formally, we define a game-theoretic centrality measure as a pair  $(\psi, \phi)$  consisting of a representation function  $\psi$  and a solution concept  $\phi$ .

**Example 5.** Let us consider a game-theoretic centrality measure  $(\psi_B, \phi^{SV})$ . We say that  $\psi_B$  is a representation function since it associates a coalitional game with any graph  $G = (V, E)$ , i.e. every graph represents a cooperative game. We have  $\psi_B(G) = (V, \nu_G)$ , where  $V$  is the set of nodes and  $\nu_G : 2^V \rightarrow \mathbb{R}$  is the characteristic function. Let  $\nu_G$  be the ranking of groups of nodes in  $G$  based on group betweenness centrality. In other words,  $\psi_B$  is simply the group betweenness  $c_{gb}$  centrality for any graph. For a specific graph  $G$ , the importance of each node  $u \in V$  according to  $(\psi_B, \phi^{SV})$  is evaluated by the Shapley value of the game  $\psi_B(G)$ , i.e.  $\phi^{SV}(\nu_G)$ . Since we started off with a well-known centrality measure (betweenness centrality) and applied a game-theoretic solution concept to it (the Shapley value), we call the resulting centrality measure a game-theoretic extension of betweenness centrality, or the Shapley-value based betweenness centrality.

The algorithms presented in the next section are designed for the generalization of the following two game-theoretic centralities.

**Definition 26 (Shapley-value based degree centrality).** Shapley-value based degree centrality is a pair  $(\psi_D, \phi^{SV})$  consisting of a representation function  $\psi_D(G) = c_{gd}$  that maps each graph to the game with group degree centrality and a Shapley value solution concept  $\phi^{SV}$ .

**Definition 27 (Shapley-value based closeness centrality).** Shapley-value based closeness centrality is a pair  $(\psi_C, \phi^{SV})$  consisting of a representation function  $\psi_C(G) = c_{gc}$  that maps each graph to the game with group closeness centrality and a Shapley value solution concept  $\phi^{SV}$ .

---

<sup>2</sup> The graphs can be simple, weighted, or directed, it will depend on the context.

### 4.3. The Algorithms

In this section, we present five characteristic function formulations, each designed to convey a specific centrality notion. These games are a generalizations of Shapley-value based degree centrality and Shapley-value based closeness centrality.

We need to clarify one important aspect of group centralities. From the Definitions 21 and 23 we see that for each graph  $G = (V, E)$  we have  $c_{gd}(V) = c_{gc}(V) = 0$ . However, in the four games  $g_1, g_2, g_3$  and  $g_5$  being under consideration in this chapter we slightly modify Definition 21 and assume that for each coalition  $C \subseteq V$  group degree centrality is:  $c_{gd}(C) + |C|$ . This assumptions were made in order to catch the intuition that the value of coalition  $C$  is in fact the size of the 'sphere of influence' made by this coalition on the network. The coalition  $C$  is expected to make influence on all its members. So, the five games analyzed in this section are:

- $g_1$  In this game the value of coalition  $C$  is a function of its own size and of the number of nodes that are immediately reachable from  $C$ . It is simply the Shapley-value based degree centrality.
- $g_2$  In this game the value of coalition  $C$  is a function of its own size and of the number of nodes that are immediately reachable in at least  $k$  different ways from  $C$ . This game is inspired by Bikhchandani et al. [1992] and is an instance of the general threshold model introduced by Kempe et al. [2005]. It has a natural interpretation: an agent "becomes influenced" (with ideas, information, marketing message, etc.) only if at least  $k$  of his neighbors have already become influenced.
- $g_3$  This game concerns weighted graphs (unlike  $g_1$  and  $g_2$ ). Here, the value of coalition  $C$  depends on its size and on the set of all nodes within a cutoff distance of  $C$ , as measured by the shortest path lengths on the weighted graph.
- $g_4$  This game generalizes  $g_3$  by allowing the value of  $C$  to be an arbitrary non-increasing function  $f(\cdot)$  of the distance between  $C$  and the other nodes in the network. The intuition here is that the coalition has more influence on closer nodes than on those further away—a property that cannot be expressed with the standard closeness centrality. Thus,  $g_4$  is the Shapley-value based closeness centrality.
- $g_5$  The last game is an extension of  $g_2$  to the case of weighted networks. Here, the value of  $C$  depends on the adjacent nodes that are connected to the coalition with weighted edges whose sum exceeds a given threshold  $w_{cutoff}$  (recall that in  $g_2$  this threshold is defined simply by the integer  $k$ ). Whereas in  $g_3$  and  $g_4$  weights on edges are interpreted as distance, in  $g_5$  they should be interpreted as a *power of influence*.

The relationships among all five games are graphically presented in Figure 4.2.

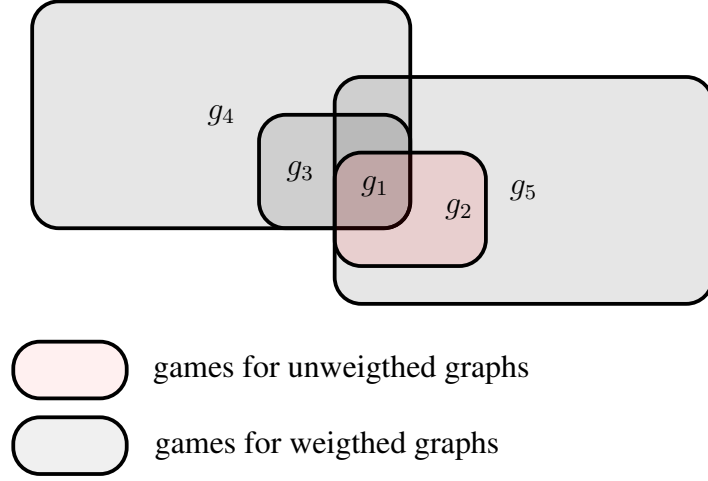


Figure 4.2: Euler diagram showing the relationships among all five games considered in this chapter. Specifically, game  $g_2$  generalizes  $g_1$ ;  $g_3$  generalizes  $g_1$  and is further generalized by  $g_4$ ;  $g_5$  generalizes  $g_2$ . Finally, we note that there are certain instances of games that can be represented as  $g_3$ ,  $g_4$  and  $g_5$ .

The computation of the Shapley value for each of the above five games (see Table 4.1 for an overview) is the main focus of the paper by Michalak et al. [2013a]. In this dissertation we skip the technical part of constructing algorithms, and instead of this, for each game we present the extensive evaluation.

#### 4.3.1. Shapley Value-Based Degree Centrality

Let  $G(V, E)$  be an unweighted, undirected network. The characteristic function is defined as:

$$\psi_1(G)(C) = \nu_1(C) = \nu_{gd}(C) + |C|$$

.

The above game was applied by Suri and Narahari [2010] to find out influential nodes in social networks and it was shown to deliver very promising results concerning the target set selection problem (see Kempe et al. [2003]). It is therefore desired to compute the Shapley values of all nodes for this game. We shall now present an exact algorithm for this computation rather than obtaining results through Monte Carlo simulation as was done by Suri and Narahari.

In more detail, to evaluate the Shapley value of node  $v$ , consider all possible permutations of the nodes in which  $v$  would make a positive marginal contribution to the coalition of nodes occurring before itself. Let the set of nodes occurring before node  $v$  in a random permutation of nodes be denoted  $P_\pi(v)$ . Let the neighbours of node  $v$  in the graph  $G(V, E)$  be denoted  $N(v)$  and the degree of node  $v$  be denoted  $deg(v)$ .

Game	Graph	Value of a coalition $C$ , i.e., $\nu(C)$	Complexity	Accuracy
$g_1$	$UW$	$\nu(C)$ is the number of nodes in $C$ and those immediately reachable from $C$	$O( V  +  E )$	exact
$g_2$	$UW$	$\nu(C)$ is the number of nodes in $C$ and those immediately reachable from $C$ , but via at least $k$ different edges	$O( V  +  E )$	exact
$g_3$	$W$	$\nu(C)$ is the number of nodes in $C$ and those not further than $d_{cutoff}$ away	$O( V  E  +  V ^2 \log  V )$	exact
$g_4$	$W$	$\nu(C)$ is the sum of $f(\cdot)$ 's — the non-increasing functions of the distance between $C$ and other nodes	$O( V  E  +  V ^2 \log  V )$	exact
$g_5$	$W$	$\nu(C)$ is the number of nodes in $C$ and those directly connected to $C$ via edges which sum of weights exceeds $W_{cutoff}$	$O( V  E )$	approx. ~ 5-10%

Table 4.1: Games considered in this dissertation and our results ( $UW$  denotes unweighted graphs and  $W$  weighted).

The key question to ask is: what is the necessary and sufficient condition for node  $v$  to marginally contribute node  $u \in N(v) \cup \{v\}$  to  $\text{fringe}(P_\pi(v))$ ? Clearly, this happens if and only if neither  $u$  nor any of its neighbours are present in  $P_\pi(v)$ . Formally,  $(N(u) \cup \{v\}) \cap P_\pi(v) = \emptyset$ .

Now we are going to show that the above condition holds with probability  $\frac{1}{1+deg(u)}$ .

**Proposition 1.** *The probability that in a random permutation none of the vertices from  $N(u) \cup \{u\}$  occurs before  $v$ , where  $v$  and  $u$  are neighbours, is  $\frac{1}{1+deg(u)}$ .*

*Proof.* We need to count the number of permutations that satisfy:

$$\forall_{n \in (N(v) \cup \{u\})} \pi(n) > \pi(v). \quad (4.1)$$

To this end:

- Let us choose  $|N(u) \cup \{u\}|$  positions in the sequence of all elements from  $V$ . We can do this in  $\binom{|V|}{1+deg(u)}$  ways.
- Then, in the last  $deg(u)$  chosen positions, place all elements from  $(N(u) \cup \{u\}) \setminus \{v\}$ . Directly before these, place the element  $v$ . The number of such line-ups is  $(deg(u))!$ .
- The remaining elements can be arranged in  $(|V| - (1 + deg(u)))!$  different ways.

All in all, the number of permutations satisfying condition (4.1) is:

$$\binom{|V|}{1+deg(u)} (deg(u))! (|V| - (1 + deg(u)))! = \frac{|V|!}{1+deg(u)};$$

thus, the probability that one of such permutations is randomly chosen is  $\frac{1}{1+deg(u)}$ . ■

---

**Algorithm 1:** Computing the Shapley value for Game 1

---

**Input:** Unweighted graph  $G(V, E)$

**Output:** Shapley values of all nodes in  $V$  for game  $g_1$

```
1 foreach  $v \in V$  do
2    $\phi_v^{SV} = \frac{1}{1+deg(v)}$ ;
3   foreach  $u \in N(v)$  do
4      $\phi_v^{SV} + = \frac{1}{1+deg(u)}$ ;
5   end
6 end
7 return  $\phi^{SV}$ ;
```

---

It is possible to derive some intuition from the above algorithm. If a node has a high degree, the number of terms in its Shapley value summation above is also high. But the terms themselves will be inversely related to the degree of neighboring nodes. This gives the intuition that a node will have high centrality not only when its degree is high, but also whenever its degree tends to be higher in comparison to the degree of its neighboring nodes. In other words, *power comes from being connected to those who are powerless*, a fact that is well-recognized by the centrality literature [Bonacich, 1987]. Following the same reasoning, we can also easily predict how dynamic changes to the network, such as adding or removing an edge, would influence the Shapley value.<sup>3</sup> Adding an edge between a powerful and a powerless node will add even more power to the former and will decrease the power of the latter. Naturally, removing an edge would have the reverse effect.

Algorithm 1 cycles through all nodes and through their neighbours, so its running time is  $O(|V| + |E|)$ .

Finally, we note that Algorithm 1 can be adopted to directed graphs with a couple of simple modifications. Specifically, in order to capture how many nodes we can access a given node from, the degree of a node should be replaced with indegree. Furthermore, a set of neighbours of a given node  $v$  should consist of those nodes to which an edge is directed from  $v$ .

#### 4.3.2. Shapley Value-Based $k$ -influence Degree Centrality

We now consider a more general game formulation for an unweighted graph  $G(V, E)$ , where the value of a coalition includes the number of agents that are either in the coalition or are adjacent to at least  $k$  agents who are in the coalition. Formally, we consider game  $g_2$  characterised by  $\nu_2 : 2^{V(G)} \rightarrow \mathbb{R}$ , where

---

<sup>3</sup> Many real-life networks are in fact dynamic and the challenge of developing fast streaming algorithms has recently attracted considerable attention in the literature [Lee et al., 2012].

---

**Algorithm 2:** Computing the Shapley value for Game 2

---

**Input:** Unweighted graph  $G(V, E)$ , positive integer  $k$

**Output:** Shapley value of all nodes in  $V$  for game  $g_2$

```
1 foreach  $v \in V$  do
2    $\phi_v^{SV} = \min(1, \frac{k}{1+deg(v)});$ 
3   foreach  $u \in N(v)$  do
4      $\phi_v^{SV} += \max(0, \frac{deg(u)-k+1}{deg(u)(1+deg(u))});$ 
5   end
6 end
7 return  $\phi^{SV};$ 
```

---

$$\psi_2(G)(C) = \nu_2(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ |\{v : v \in C \text{ (or) } |N(v) \cap C| \geq k\}| & \text{otherwise.} \end{cases}$$

The second game is an instance of the General Threshold Model that has been widely studied in the literature [Kempe et al., 2005, Granovetter, 1978]. Intuitively, in this model each node can become active if a monotone activation function reaches some threshold. The instance of this problem has been proposed by Goyal et al. [2010], where the authors introduced a method of learning influence probabilities in social networks (from users' action logs). However, in many realistic situations much less information is available about a network so it is not possible to assess specific probabilities with which individual nodes become active. Consequently, much simpler models are studied. Bikhchandani et al. [1992], for instance, “consider a teenager deciding whether or not to try drugs. A strong motivation for trying out drugs is the fact that friends are doing so. Conversely, seeing friends reject drugs could help persuade the teenager to stay clean”. This situation is modelled by the second game; the threshold for each node is  $k$  and the activation function is  $f(S) = |S|$ . Another example is viral marketing or innovation diffusion analysis. Again, in this application, it is often assumed that an agent will “be influenced” only if at least  $k$  of his neighbors have already been convinced by Valente [1996]. Note that this game reduces to game  $g_1$  for  $k = 1$ .

Although this game is a generalization of game  $g_1$ , it can still be solved to obtain the Shapley values of all nodes in  $O(|V| + |E|)$  time, as formalised by Algorithm 2.

An even more general formulation of the game is possible by allowing  $k$  to be a function of the agent, i.e., each node  $v \in V$  is assigned its own unique attribute  $k(v)$ . This translates to an application of the form: agent  $i$  is convinced if and only if at least  $k_v$  of his neighbors are convinced, which is a frequently used model in the literature [Valente, 1996].

The above argument does not use the fact that  $k$  is constant across all nodes. So this generalized formulation can be solved by a simple modification to the original Shapley value expression:

$$SV(v) = \frac{k(v)}{1 + \deg(v)} + \sum_{u \in N(v)} \frac{1 + \deg(u) - k(u)}{\deg(u)(1 + \deg(u))}.$$

The above equation (which is also implementable in  $O(|V| + |E|)$  time) assumes that  $k(v) \leq 1 + \deg(v)$  for all nodes  $v \in V$ . This condition can be assumed without loss of generality because all cases can still be modelled (we set  $k(v) = 1 + \deg(v)$  for the extreme case where node  $v$  is never convinced no matter how many of its neighbors are already convinced).

Finally, we note that Algorithm 2 can be adapted to a case of directed graphs along the same lines as Algorithm 1.

#### 4.3.3. Shapley Value-Based $d_{cutoff}$ -distance Degree Centrality

Hitherto, our games have been confined to unweighted networks. But in many applications, it is necessary to model real-world networks as weighted graphs. For example, in the co-authorship network mentioned in the introduction, each edge is often assigned a weight proportional to the number of joint publications the corresponding authors have produced [Newman, 2001].

This subsection extends game  $g_1$  to the case of weighted networks. Whereas game  $g_1$  equates  $\nu(C)$  to the number of nodes located within one hop of some node in  $C$ , our formulation in this subsection equates  $\nu(C)$  to the number of nodes located within a distance  $d_{cutoff}$  of some node in  $C$ . Here, distance between two nodes is measured as the length of the shortest path between the nodes in the given weighted graph  $G = (V, E, \lambda)$  (see Definition 13)

Formally, we define game  $g_3$ , where for each coalition  $C \subseteq V(G)$ ,

$$\psi_3(G)(C) = \nu_3(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ |\{v : \text{dist}(C, v) \leq d_{cutoff}\}| & \text{otherwise.} \end{cases}$$

Clearly,  $g_3$  can be used in all the settings where  $g_1$  is applicable; for instance, in the diffusion of information in social networks or to analyse research collaboration networks [Suri and Narahari, 2010]. Moreover, as a more general game,  $g_3$  provides additional modelling opportunities. For instance, Suri and Narahari suggest that a “*more intelligent*” way for sieving nodes in the neighbourhood would improve their algorithm for solving the target selection problem (*top-k* problem). Now,  $g_3$  allows us to define a different *cutoff* distance for each node in Suri and Narahari’s setting. Furthermore,  $g_3$  is a specific case of the more general model  $g_4$  which will be discussed in next subsection.

---

**Algorithm 3:** Computing the Shapley value for Game 3

---

**Input:** Weighted graph  $G = (V, E, \lambda)$ ,  $d_{cutoff} > 0$

**Output:** Shapley value of all nodes in  $G$  for game  $g_3$

```
1 foreach  $v \in V(G)$  do
2   DistanceVector  $D = \text{Dijkstra}(v, G)$ ;
3    $\text{extNeighbors}(v) = \emptyset$ ;  $\text{extDegree}(v) = 0$ ;
4   foreach  $u \in V(G)$  such that  $u \neq v$  do
5     if  $D(u) \leq d_{cutoff}$  then
6        $\text{extNeighbors}(v).push(u)$ ;
7        $\text{extDegree}(v)++$ ;
8     end
9   end
10 end
11 foreach  $v \in V(G)$  do
12    $\phi_v^{SV} = \frac{1}{1+\text{extDegree}(v)}$ ;
13   foreach  $u \in \text{extNeighbors}(v)$  do
14      $\phi_v^{SV} += \frac{1}{1+\text{extDegree}(u)}$ ;
15   end
16 end
17 return  $\phi^{SV}$ ;
```

---

Before we propose an algorithm, we need to introduce some extra notation. Define the *extended neighborhood*  $N(v, d_{cutoff}) = \{u \neq v : \text{dist}(v, u) \leq d_{cutoff}\}$ , i.e., the set of all nodes whose distance from  $v$  is at most  $d_{cutoff}$ . Denote the size of  $N_G(v, d_{cutoff})$  by  $\text{deg}(v, d_{cutoff})$ .

Algorithm 3 works as follows: for each node  $v$  in the network  $G(V, E)$ , the extended neighborhood  $N(v, d_{cutoff})$  and its size  $\text{deg}(v, d_{cutoff})$  are first computed using Dijkstra's algorithm in  $O(|E| + |V|\log|V|)$  time [Cormen, 2001]. The results are then used to compute Shapley value, which takes maximum time  $O(|V|^2)$ . In practice this step runs much faster because the worst case situation only occurs when every node is reachable from every other node within  $d_{cutoff}$ . Overall the complexity is  $O(|V||E| + |V|^2\log|V|)$ .

Furthermore, to deal with directed graphs we need to redefine the notion of *extDegree* and *extNeighbors* for a given node  $u$  in Algorithm 3. The former will be the number of vertices from which the distance to  $u$  is smaller than, or equal to,  $d_{cutoff}$ . The latter will be the set of nodes whose distance from  $u$  is at most  $d_{cutoff}$ .

Finally, we make the observation that the the above algorithm does not depend on  $d_{cutoff}$  being constant across all nodes. Indeed, each node  $v \in V(G)$  may be assigned its own unique value  $d_{cutoff}(v)$ , where  $\nu(C)$  would be the number of agents  $v$  who are within a distance  $d_{cutoff}(v)$  from  $C$ . For this case, the above proof gives:



$$\phi_v^{SV} = \sum_{\substack{u: \text{dist}(v,u) \\ \leq d_{\text{cutoff}}(u)}} \frac{1}{1 + \text{deg}(u, d_{\text{cutoff}}(u))}.$$

#### 4.3.4. Shapley Value-Based Closeness Centrality

This subsection further generalizes game  $g_3$ , again taking motivation from real-life network problems. In game  $g_3$ , all agents at distances  $d_{\text{agent}} \leq d_{\text{cutoff}}$  contributed equally to the value of a coalition. However, this assumption may not always hold true because in some applications we intuitively expect agents closer to a coalition to contribute more to its value. For instance, we expect a Facebook user to exert more influence over his immediate circle of friends than over “friends of friends”, even though both may satisfy the  $d_{\text{cutoff}}$  criterion. Similarly, we expect a virus-affected computer to infect a neighboring computer more quickly than a computer two hops away.

In general, we expect that an agent at distance  $d$  from a coalition would contribute  $f(d)$  to its value, where  $f(\cdot)$  is a positive valued decreasing function of its argument. More formally, we define game  $g_4$ , where the value of a coalition  $C$  is given by:

$$\psi_4(G)(C) = \nu_4(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ \sum_{v \in V} f(\text{dist}(C, v)) & \text{otherwise.} \end{cases}$$

In the case of Shapley value-based closeness centrality, the key question to ask is: what is the expected value of the marginal contribution of  $v$  through node  $u \neq v$  to the value of coalition  $P_\pi(v)$ ? Let this marginal contribution be denoted  $MC(v, u)$ . Clearly:

$$MC(v, u) = \begin{cases} 0 & \text{if } \text{dist}(v, u) \geq \text{dist}(u, P_\pi(v)) \\ f(\text{dist}(v, u)) - f(\text{dist}(u, P_\pi(v))) & \text{otherwise.} \end{cases}$$

Let  $D_u = \{d_1, d_2 \dots d_{|V|-1}\}$  be the distances of node  $u$  from all other nodes in the network, sorted in increasing order. Let the nodes corresponding to these distances be  $\{w_1, w_2 \dots w_{|V|-1}\}$ , respectively. Let  $k_{vu} + 1$  be the number of nodes (out of these  $|V| - 1$ ) whose distances to  $u$  are  $\leq \text{distance}(v, u)$ . Let  $w_{k_{vu}+1} = v$  (i.e., among all nodes that have the same distance from  $u$  as  $v$ ,  $v$  is placed last in the increasing order).

We use *literal*  $w_i$  to mean  $w_i \in P_\pi(v)$  and the literal  $\overline{w_i}$  to mean  $w_i \notin P_\pi(v)$ . Define a sequence of boolean variables  $p_k = \overline{u} \wedge \overline{w_1} \wedge \overline{w_2} \wedge \dots \wedge \overline{w_k}$  for each  $0 \leq k \leq |V| - 1$ . Finally denote expressions of the form  $MC(v, u|F)$  to mean the marginal contribution of  $v$  to  $P_\pi(v)$  through  $u$  given that the coalition  $P_\pi(v)$  satisfies the boolean expression  $F$ .

$$\begin{aligned}
MC(v, u|p_{k_{vu}+1} \wedge w_{k_{vu}+2}) &= f(d_{k_{vu}+1}) - f(d_{k_{vu}+2}), \\
MC(v, u|p_{k_{vu}+2} \wedge w_{k_{vu}+3}) &= f(d_{k_{vu}+1}) - f(d_{k_{vu}+3}), \\
&\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
MC(v, u|p_{|V|-2} \wedge w_{|V|-1}) &= f(d_{k_{vu}+1}) - f(d_{|V|-1}), \\
MC(v, u|p_{|V|-1}) &= f(d_{k_{vu}+1}).
\end{aligned}$$

With this notation, we obtain expressions for  $MC(v, u)$  by splitting over the above *mutually exclusive* and *exhaustive* (i.e., covering all possible non-zero marginal contributions) cases.

Now, we need to determine the probability of  $\Pr(p_k \wedge w_{k+1})$ .

**Proposition 2.** *The probability that in a random permutation none of the nodes from  $\{v_j, w_1, \dots, w_k\}$  occur before  $v_i$  and the node  $w_{k+1}$  occurs before  $v_i$  is  $\frac{1}{(k+1)(k+2)}$ .*

*Proof.* Let us count the number of permutations that satisfy:

$$\forall_{v \in \{v_j, w_1, \dots, w_k\}} \pi(v) < \pi(v) \wedge \pi(v < \pi(w_{k+1})). \quad (4.2)$$

To this end:

- Let us choose  $|\{v_j, w_1, \dots, w_k\} \cup \{u\} \cup \{w_{k+1}\}|$  positions in the sequence of all elements from  $V$ . We can do this in  $\binom{|V|}{k+3}$  ways.
- Then, in the last  $k+1$  chosen positions, we place all elements from  $\{v_j, w_1, \dots, w_k\}$ . Directly before these, we place the element  $v$ , and then vertex  $w_{k+1}$ . The number of such line-ups is  $(k+1)!$ .
- The remaining elements can be arranged in  $(|V| - (k+3))!$  different ways.

Thus, the number of permutations satisfying (4.2) is:

$$\binom{|V|}{k+3} (k+1)! (|V| - (k+3))! = \frac{|V|!}{(k+1)(k+2)},$$

and the probability that one of such permutations is randomly chosen is  $\frac{1}{(k+1)(k+2)}$ . ■

With the above proposition we find that:

$$\Pr(p_k \wedge w_{k+1}) \frac{1}{(k+1)(k+2)} \quad \forall \quad 1 + k_{vu} \leq k \leq |V| - 2.$$

Using the  $MC(v, u)$  equations and the probabilities  $\Pr(p_k \wedge w_{k+1})$ :

---

**Algorithm 4:** Computing the Shapley value for Game 4
 

---

**Input:** Weighted graph  $G = (V, E, \lambda)$ , function  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$

**Output:** Shapley value of all nodes in  $G$  for game  $g_4$

```

1 Initialise:  $\forall v \in V(G)$  set  $\phi_v^{SV} = 0$ ;
2 foreach  $v \in V(G)$  do
3   [Distances  $D$ , Nodes  $w$ ] = Dijkstra( $v, G$ );
4   sum = 0; index =  $|V|-1$ ; prevDistance = -1, prevSV = -1;
5   while index > 0 do
6     if  $D(\text{index}) == \text{prevDistance}$  then
7       | currSV = prevSV;
8     else
9       | currSV =  $\frac{f(D(\text{index}))}{1+\text{index}} - \text{sum}$ ;
10    end
11    node =  $w(\text{index})$ ;
12     $\phi_{\text{node}}^{SV} + = \text{currSV}$ ;
13    sum +=  $\frac{f(D(\text{index}))}{\text{index}(1+\text{index})}$ ;
14    prevDistance =  $D(\text{index})$ , prevSV = currSV;
15    index--;
16  end
17   $\phi_v^{SV} + = f(0) - \text{sum}$ ;
18 end
19 return  $\phi^{SV}$ ;

```

---

$$\begin{aligned}
 E[MC(v, u)] &= \left[ \sum_{k=1+k_{vu}}^{|V|-2} \frac{f(\text{distance}(v, u)) - f(d_{k+1})}{(k+1)(k+2)} \right] + \frac{f(\text{distance}(v, u))}{|V|} \\
 &= \frac{f(\text{distance}(v, u))}{k_{vu} + 2} - \sum_{k=k_{vu}+1}^{|V|-2} \frac{f(d_{k+1})}{(k+1)(k+2)}.
 \end{aligned}$$

In Algorithm 4 we use  $D_v = \{d_1, d_2 \dots d_{|V|-1}\}$  as a vector of distances from node  $v$  to all other nodes in the network, sorted in increasing order. Also, the vector  $w = \{w_1, w_2 \dots w_{|V|-1}\}$  represents nodes corresponding to these distances, respectively. For each vertex  $v$ , a vector of distances to every other vertex is first computed using Dijkstra's algorithm [Cormen, 2001]. This yields a vector  $D_v$  that is already sorted in increasing order. This vector is then traversed in reverse to compute the Shapley value of the appropriate node  $w$ . After the traversal, the Shapley value of  $v$  itself is updated. This process is repeated for all nodes  $v$  so that at the end of the algorithm, the Shapley value is computed exactly in  $O(|V||E| + |V|^2 \log |V|)$  time.

Our final observation is that Algorithm 4 works also for directed graphs as long as we use the appropriate version of Dijkstra’s algorithm (see, e.g., Cormen [2001]).

#### 4.3.5. Shapley Value-Based $W_{cutoff}$ -influence Degree Centrality

In this subsection, we generalize game  $g_2$  for the case of weighted networks. Given a positive weighted network  $G = (V, E, \lambda)$  and a value  $W_{cutoff}(v)$  for every node  $v \in V$ , we first define  $\lambda(v, C) = \sum_{u \in C} \lambda(v, u)$  for every coalition  $C$ , where  $\lambda(v, u)$  is the weight of the edge between nodes  $v$  and  $u$  (or 0 if there is no such edge). With this notation, we define game  $g_5$  by the characteristic function:

$$\nu_5(C) = \begin{cases} 0 & \text{if } C = \emptyset \\ \text{size}(\{v : v \in C \text{ (or) } \lambda(v, C) \geq W_{cutoff}(v)\}) & \text{otherwise.} \end{cases}$$

The formulation above has applications in, for instance, the analysis of information diffusion, adoption of innovations, and viral marketing. Indeed, many cascade models of such phenomena on weighted graphs have been proposed [Granovetter, 1978, Kempe et al., 2003, Young, 2005] which work by assuming that an agent will change state from “inactive” to “active” if and only if the sum of the weights to all active neighbors is at least equal to an agent-specific cutoff.

Although we have not been able to come up with an exact formula for the Shapley value in this game<sup>4</sup>, our analysis yields an approximate formula which was found to be accurate in practice.

We will need the following additional notation:

- let the weights of edges between  $v$  and each of the nodes in  $N(v)$  be  $\lambda_v = \{\lambda(v, u), \lambda_1, \lambda_2 \dots \lambda_{deg(v)-1}\}$  in that order;
- let  $\alpha_v$  be the sum of all the weights in  $\lambda_v$  and  $\beta_v$  be the sum of the squares of all the weights in  $\lambda_v$ ; and finally
- let  $X_t^{vu}$  be the sum of a  $t$ -subset of  $\lambda_v \setminus \{\lambda(v, u)\}$  drawn uniformly at random from the set of all such possible  $t$ -subsets.

While in each graph it holds that  $\sum_{v \in V} deg(v) \leq 2|E|$ , Algorithm 5 implements an  $O(|V| + \sum_{v \in V} \sum_{u \in N(v)} deg(u)) \leq O(|V| + |V||E|) = O(|V||E|)$  solution to compute the Shapley value for all agents in game  $g_5$  using the above approximation. Furthermore, we make the following observation: for small  $deg(u)$ , one might as well use the brute force computation to determine Shapley value in  $O(2^{deg(u)-1})$  time.

---

<sup>4</sup> Computing the Shapley value for this game involves determining whether the sum of weights on specific edges, adjacent to a random coalition, exceeds the threshold. This problem seems to be at least as hard as computing the Shapley value in weighted voting games, which is #P-Complete [Elkind et al., 2009b].

---

**Algorithm 5:** Computing the Shapley value for Game 5

---

**Input:** Weighted network  $G = (V, E, \lambda)$ , cutoffs  $W_{cutoff}(v)$  for each  $v \in V$

**Output:** Shapley value of all nodes in  $G$  for game  $g_5$

```
1 foreach  $v \in V$  do
2   | compute and store  $\alpha_i$  and  $\beta_i$ ;
3 end
4 foreach  $v \in V$  do
5   |  $\phi_v^{SV} = 0$ ;
6   | foreach  $m$  in 0 to  $\deg(v)$  do
7     | compute  $\mu = \mu(X_m^{vv})$ ,  $\sigma = \sigma(X_m^{vv})$ ,  $p = \Pr\{\mathcal{N}(\mu, \sigma^2) < W_{cutoff}(v)\}$ ;
8     |  $\phi_v^{SV} += \frac{p}{1+\deg(v)}$ ;
9   | end
10  | foreach  $u \in N(v)$  do
11    |  $p = 0$ ;
12    | foreach  $m$  in 0 to  $\deg(u) - 1$  do
13      | compute  $\mu = \mu(X_m^{vu})$ ,  $\sigma = \sigma(X_m^{vu})$  and  $z = Z_m^{vu}$ ;
14      |  $p += z \frac{\deg(u)-m}{\deg(u)(\deg(u)+1)}$ ;
15    | end
16    |  $\phi_v^{SV} += p$ ;
17  | end
18 end
19 return  $\phi^{SV}$ ;
```

---

As far as directed graphs are concerned, in all calculations in Algorithm 5 we have to consider the indegree of a node instead of degree. Furthermore, the set of neighbours of a node  $u$  should be defined as the set of nodes  $v$  connected with directed edge  $(u, v)$ .

## 4.4. Simulations

In this section we evaluate the time performance of our exact algorithms for games  $g_1$  to  $g_4$  and our approximation algorithm for game  $g_5$ . In more detail, we compare our exact algorithms to the method of approximating the Shapley value via Monte Carlo sampling which has been the only feasible approach to compute game-theoretic network centrality available to date in the literature. In this section we evaluate the time performance of our exact algorithms for games  $g_1$  to  $g_4$  and our approximation algorithm for game  $g_5$ . In more detail, we compare our exact algorithms to the method of approximating the Shapley value via Monte Carlo sampling which has been the only feasible approach to compute game-theoretic network centrality available to date in the literature. First, we provide a detailed description of the simulation setup; then, we present data sets and the simulation results. In the remainder of this section we first we

comment on Monte Carlo sampling. Next, we provide a detailed description of the simulation setup. Finally, we present data sets and the simulation results.

#### 4.4.1. Approximation Methods for the Shapley Value

In this section, we discuss different approximation techniques to compute the Shapley value from the literature. Generally speaking, they can be divided into three groups—each referring to a specific subclass of coalitional games under consideration:

1. First, let us consider the method proposed by Fatima et al. [2007] and elaborated further by Fatima et al. [2008]. This approach concerns *weighted voting games*. In these games, each player has a certain number of votes (or in other words, a weight). A coalition is “winning” if the number of votes in this coalition exceeds some specific threshold, or “losing” otherwise. Fatima et al. propose the following method to approximate the Shapley value in weighted voting games. Instead of finding marginal contributions of players to all  $2^n$  coalitions, the authors consider only  $n$  randomly-selected coalitions, one of each size (i.e., from 1 to  $n$ ). Only for these  $n$  coalitions are the player’s expected marginal contributions calculated and the average of these contributions yields an approximation of the Shapley value. Whereas Fatima et al. method is certainly attractive, it is only applicable to games in which the value of a coalition depends on the sum of associated weights being in some bounds. This is not the case for our games  $g_1$  to  $g_4$ .<sup>5</sup>
2. Another method was proposed by Bachrach et al. [2008] in the context of *simple coalitional games*<sup>6</sup> in which the characteristic function is binary—i.e., each coalition has a value of either zero or one. For these games, Bachrach et al. extend the approach suggested by Mann and Shapley 1960 and provide more rigorous statistical analysis. In particular, Mann and Shapley described the Monte Carlo simulations to estimate the Shapley value from a random sample of coalitions. Bachrach et al. use this technique to compute the Banzhaf power index and then they suggested using a random sample of permutations of all players in order to compute the Shapley-Shubik index for simple coalitional games.<sup>7</sup> The computation of the confidence interval, which is crucial in such an approach, hinges upon the binary form of the characteristic function for simple coalitional games. This method is more general than the one proposed by Fatima et al. [2007]—as weighted voting games are a subset of simple coalitional games—but still it cannot be effectively used for our games  $g_1$  to  $g_4$ , where the characteristic functions are not binary.

---

<sup>5</sup> Recall that our approximation algorithm for  $g_5$  builds upon Fatima et al. method. This is because in this game the marginal contribution of each node depends on the weights assigned to its incident edges.

<sup>6</sup> Note that weighted voting games are simple coalitional games.

<sup>7</sup> The Shapley-Shubik index is a very well-known application of the Shapley value that evaluates the power of individuals in voting [Shapley and Shubik, 1954].

---

**Algorithm 6:** Monte Carlo method to approximate the Shapley value

---

**Input:**

◇ Characteristic function  $v$ , maximum iteration  $maxIter$

**Output:** Aproximation of Shapley value for game  $v$

```
1 for  $v \in V$  do
2    $\phi_v^{SV} = 0$  ;
3 end
4 for  $i = 1$  to  $maxIter$  do
5   shuffle( $V$ );
6   Marginal Contribution block
7      $P = \emptyset$  ;
8     for  $v \in V$  do
9        $\phi_v^{SV} += v(P \cup \{v\}) - v(P)$  ;
10       $P = P \cup \{v\}$  ;
11    end
12
13 end
14 for  $v \in V$  do
15    $\phi_v^{SV} = \frac{\phi_v^{SV}}{maxIter}$  ;
16 end
17 return  $\phi_v^{SV}$  ;
```

---

3. Unlike the first two methods, the last method described by Castro et al. [2009] can be efficiently applied to all coalitional games in characteristic function game form, assuming that the worth of every coalition can be computed in polynomial time. Here, approximating the Shapley value involves generating permutations of all players and computing the marginal contribution of each player to the set of players occurring before it. The solution precision increases (statistically) with every new permutation analysed. Furthermore, the authors show how to estimate the appropriate size of a permutation sample in order to guarantee a low error. Given its broad applicability, this method is used in our simulations as a comparison benchmark.

#### 4.4.2. Simulation Setup

In more detail, in a preliminary step, we test what is the maximum number of Monte Carlo iterations that can be performed in a reasonable time for any given game. This maximum number of iterations, denoted  $maxIter$ , becomes an input to Algorithm 6 for Monte Carlo sampling. In this algorithm, in each one of the  $maxIter$  iterations, a random permutation of all nodes is generated. Then, using a characteristic function from the set  $\nu \in \{\nu_1, \nu_2, \nu_3, \nu_4, \nu_5\}$ , it calculates the marginal contribution of each node to the set  $P$  of nodes occurring before a

given node in a random permutation.<sup>8</sup> Finally, the algorithm divides the aggregated sum of all contributions for each node by the number of iterations performed. The time complexity of this algorithm is  $O(\maxIter * con)$ , where  $con$  denotes the number of operations necessary for computing the *Marginal Contribution* block. This block is specifically tailored to the particular form of the characteristic function of each of the games  $g_1$  to  $g_5$ . In particular, for game  $g_1$  (see Algorithm 6), it is constructed as follows. Recall that, in this game, node  $v_i$  makes a positive contribution to coalition  $P$  through itself and through some adjacent node  $u$  under two conditions. Firstly, neither  $v_i$  nor  $u$  are in  $P$ . Secondly, there is no edge from  $P$  to  $v_i$  or  $u$ . To check for these conditions in Algorithm 6 we store those nodes that have already contributed to the value of coalition  $P$  in an array called: *Counted*. For each node  $v_i$ , the algorithm iterates through the set of its neighbours and for each adjacent node it checks whether this adjacent node is counted in the array *Counted*. If not, the marginal contribution of the node  $v_i$  is increased by one.

---

**Algorithm 7: *Marginal Contribution* block of Algorithm 6 for  $g_2$**

---

```

Counted  $\leftarrow$  false ;
Edges  $\leftarrow$  0 ;
foreach  $v \in V(G)$  do
    foreach  $u \in N_G(v) \cup \{v\}$  do
        Edges[u]++ ;
        if !Counted[u] and ( Edges[u]  $\geq$  k[u] or  $u = v$  ) then
             $\phi_v^{SV} \leftarrow \phi_v^{SV} + 1$  ;
            Counted[u] = true ;
        end
    end
end

```

---



---

**Algorithm 8: *Marginal Contribution* block of Algorithm 6 for  $g_3$**

---

```

Counted  $\leftarrow$  false ;
foreach  $v \in V(G)$  do
    foreach  $u \in extNeighbors(v) \cup \{v\}$  do
        if !Counted[u] then
             $\phi_v^{SV} \leftarrow \phi_v^{SV} + 1$  ;
            Counted[u] = true ;
        end
    end
end

```

---

For the rest of the games considered in this chapter the *Marginal Contribution* block of Algorithm 6 takes a slightly different form. We have explained the functioning of this block for  $g_1$ . Now, we discuss this block for the remaining four games. In particular:

---

<sup>8</sup> Recall that the characteristic functions  $v_1, v_2, \dots, v_5$  correspond to games  $g_1, g_2, \dots, g_5$ , respectively.



---

**Algorithm 9: Marginal Contribution** block of Algorithm 6 for  $g_4$ 

---

```
dist  $\leftarrow$  infinity ;
foreach  $v \in V(G)$  do
    foreach  $u \in V(G)$  do
        if  $D[u] < dist[u]$  then
             $\phi_v^{SV} \leftarrow \phi_v^{SV} + f(D[u]) - f(dist[u]);$ 
             $dist[u] = D[u]$  ;
        end
    end
     $\phi_v^{SV} \leftarrow \phi_v^{SV} + f(dist[u]) - f(0);$ 
     $dist[v] = 0$  ;
end
```

---

---

**Algorithm 10: Marginal Contribution** block of Algorithm 6 for  $g_5$ 

---

```
Counted  $\leftarrow$  false ;
Weights  $\leftarrow$  0 ;
foreach  $v \in V(G)$  do
    foreach  $u \in N_G(v) \cup \{v\}$  do
        weights[u] +=  $\lambda(v, u)$ ;
        if !Counted[u] and ( weights[u]  $\geq W_{cutoff}(u)$  or  $u = v$  ) then
             $\phi_v^{SV} \leftarrow \phi_v^{SV} + 1$ ;
            Counted[u] = true ;
        end
    end
end
end
```

---

$g_2$ : Here, node  $v$  makes a positive contribution to a coalition  $P$  both through itself and through some adjacent node  $u$  also under two conditions. Firstly, neither  $v$  nor  $u$  are in  $P$ . Secondly, there is less than  $k$  edges from  $P$  to  $v$  and there is exactly  $k - 1$  edges from  $P$  to  $u$ . In order to check the first condition in Algorithm 7 we use the array *Counted*, and to check the second one, we use the array *Edges*. For each node  $v_i$ , the algorithm iterates through the set of its neighbours and for each adjacent node it checks whether this adjacent node meets these two conditions. If so, then the marginal contribution of the node  $v$  is increased by one.

$g_{3/4}$ : In *Marginal Contribution* blocks for games  $g_3$  and  $g_4$  (Algorithms 8 and 9), all the values that are dependent on the distance (*extNeighbours* and  $D$ ) are calculated using Dijkstra's algorithm and stored in memory. These pre-computations allow us to significantly speed up Monte Carlo methods. Now, in  $g_3$  node  $v$  makes a positive contribution to coalition  $P$  through itself and through some adjacent node  $u$  under two conditions. Firstly, neither  $v$  nor  $u$  are in  $P$ . Secondly, there is no edge length of  $d_{cutoff}$  from  $P$  to  $v$  or  $u$ . To check for these conditions in Algorithm 8 we again use the array *Counted*. For each node  $v$ , the

algorithm iterates through the set of its extended neighbours and for each of them it checks whether this neighbour meets the conditions. If so, the marginal contribution of the node  $v$  is increased by one. In game  $g_4$ , node  $v$  makes a positive contribution to coalition  $P$  through each node (including itself) that is closer to  $v_i$  than to  $P$ . In Algorithm 9 we use array  $Dist$  to store distances from coalition  $P$  to all nodes in the graph and array  $D$  to store all distances from  $v$  to all other nodes. For each node  $v_i$ , the algorithm iterates through all nodes in the graph, and for each node  $u$ , if the distance from  $v$  to  $u$  is smaller than from  $P$  to  $u$ , the algorithm computes the marginal contribution as  $f(D[u]) - f(Dist[u])$ . The value  $Dist[u]$  is then updated to  $D[u]$ —this is a new distance from  $P$  to  $u$ .

$g_5$ : In game  $g_5$ , which is an extension of  $g_2$  to weighted graphs, node  $v$  makes a positive contribution to coalition  $P$  (both through itself and through some adjacent node  $u$ ) under two conditions. Firstly, neither  $v$  nor  $u$  are in  $P$ . Secondly, the sum of weights on edges from  $P$  to  $v$  is less than  $W_{cutoff}(v_i)$  and the sum of weights on edges from  $P$  to  $u$  is greater than, or equal to,  $W_{cutoff}(u) - \lambda(v, u)$  and smaller than  $W_{cutoff}(v_i) + \lambda(v, u)$ . In order to check the first condition in Algorithm 10 we use the array  $Counted$ , and to check the second one, we use the array  $Weights$ . For each node  $v_i$ , the algorithm iterates through the set of its neighbours and for each adjacent node it checks whether this adjacent node meets these two conditions. If so, then the marginal contribution of the node  $v$  is increased by one.

Some details of how Algorithm 6 is applied to generate the Shapley value approximations for games  $g_1$  to  $g_4$ , for which we propose exact polynomial solutions, differ from  $g_5$ , for which we developed an approximate solution. Specifically, for games  $g_1$  to  $g_4$ :

1. We use the exact algorithm proposed in this chapter to compute the Shapley value.
2. Then, we run Monte Carlo simulations 30 times.<sup>9</sup> In every run:
  - We perform *maxIter* Monte Carlo iterations.
  - After every five iterations, we compare the approximation of the Shapley value obtained via Monte Carlo simulation with the exact Shapley value obtained with our algorithm.
  - We record the algorithm's runtime and the error, where the error is defined as the maximum discrepancy between the actual Shapley value and the Monte Carlo-based approximation of the Shapley value.
3. Finally, we compute the confidence interval using all iterations (0.95% confidence level).<sup>10</sup>

In the case of game  $g_5$  we cannot determine the exact Shapley value for larger networks. Therefore, we performed two levels of simulation: one level on small networks and one level on large networks. Specifically:

<sup>9</sup> For the purpose of comparison to our method, it suffices to use 30 iterations, as the standard errors converge significantly to indicate the magnitude of the cost of using the Monte Carlo method.

<sup>10</sup> Since for  $g_4$  each Monte Carlo iteration is relatively time consuming, we run it only once; thus, no confidence interval is generated, i.e., the third step is omitted.

1. For small networks, we generate 30 random instances of weighted complete graphs with 6 nodes (denoted  $K_6$ ) and the same number of graphs with 12 nodes (denoted  $K_{12}$ ) with weights drawn from a uniform distribution  $U(0, 1)$ . Then, for each graph and each of the two parameters  $W_{cutoff}(v) = \frac{1}{4}\alpha(v)$  and  $W_{cutoff}(v) = \frac{3}{4}\alpha(v)$ :<sup>11</sup>
  - We compute the exact Shapley value using formula (6).
  - Then, we run our approximation algorithm and determine the error in our approximation.
  - Finally, we run 2000 and 6000 Monte Carlo iterations for  $K_6$  and  $K_{12}$ , respectively.
2. For large networks, we again generate 30 random instances of weighted complete graphs, but now with 1000 nodes (we denote them  $K_{1000}$ ). Then, for each graph and each of the three parameters  $W_{cutoff}(v) = \frac{1}{4}\alpha(v)$ ,  $W_{cutoff}(v) = \frac{2}{4}\alpha(v)$ , and  $W_{cutoff}(v) = \frac{3}{4}\alpha(v)$ :
  - We run our approximation algorithm for the Shapley value.
  - Then, we run the fixed number (200000) of Monte Carlo iterations.
  - Finally, we compute how the Monte Carlo solution converges to the results of our approximation algorithm.

Having described the simulation setup, we will now discuss the data sets and, finally, the simulation results.

#### 4.4.3. Data Used in Simulations

We consider two networks that have already been well-studied in the literature. Specifically, for games  $g_1 - g_3$  we present simulations on an undirected, unweighted network representing the topology of the Western States Power Grid (WSPG).<sup>12</sup> This network (which has 4940 nodes and 6594 edges) has been studied in many contexts before (see, for instance, Watts and Strogatz [1998]) and is freely available online (see, e.g., <http://networkdata.ics.uci.edu/data.php?id=107>). For games  $g_3 - g_5$  (played on weighted networks), we used the network of astrophysics collaborations (abbreviated henceforth APhC) between Jan 1, 1995 and December 31, 1999. This network (which has 16705 nodes and 121251 edges) is also freely available online (see, e.g., <http://networkdata.ics.uci.edu/data.php?id=13>) and has been used in previous studies like Newman [2001].

#### 4.4.4. Simulation Results

The results presented in this section show that our exact algorithms are, in general, much faster than the Monte Carlo sampling, and this is the case even if we allow for generous error tolerance. Furthermore, requiring smaller Monte Carlo errors makes the Monte Carlo runtime exponentially slower than our exact solution.

<sup>11</sup> Recall that  $\alpha_j$  is the sum of all the weights in  $W_j$  as defined in Section 4.3.5.

<sup>12</sup> Note that with the distance threshold  $d_{cutoff}$  replaced with a hop threshold  $k_{cutoff}$ , game  $g_3$  can be played on an unweighted network.

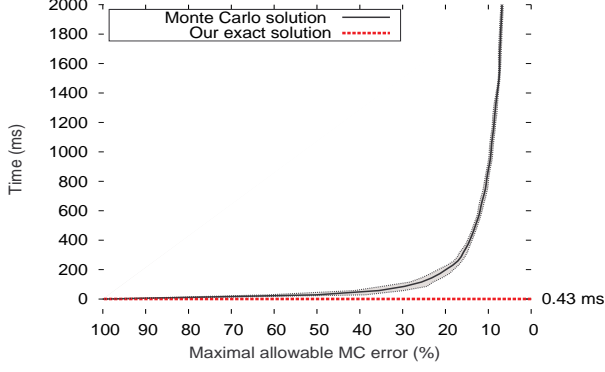


Figure 4.3:  $g_1$ , WSPG (UW)

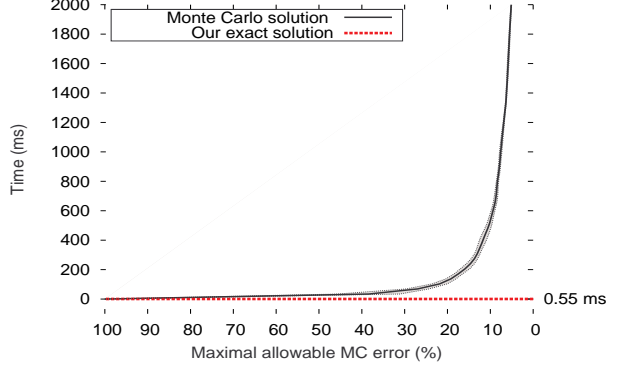


Figure 4.4:  $g_2$ ,  $k = 2$ , WSPG (UW)

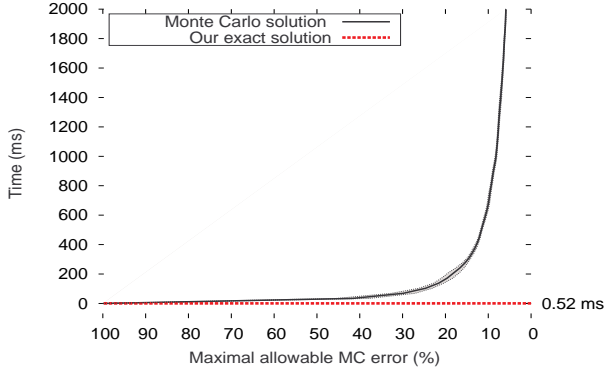


Figure 4.5:  $g_2$ ,  $k_i = \frac{deg_i}{2}$ , WSPG (UW)

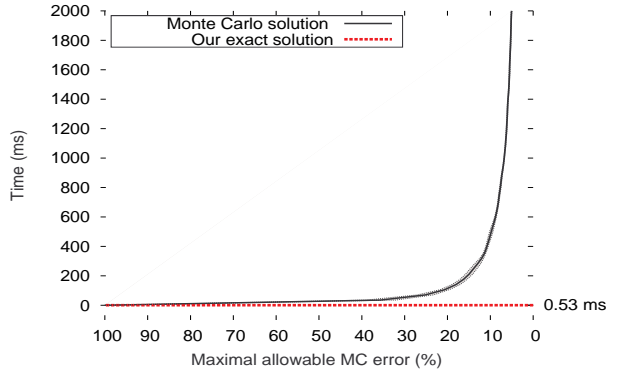


Figure 4.6:  $g_2$ ,  $k_i = \frac{3}{4} deg_i$ , WSPG (UW)

In more detail, the simulation results for game  $g_1$  are shown in Figure 4.3. The dotted line shows the performance of our exact algorithm which needs  $0.43ms$  to compute the Shapley value. In contrast, generating any reasonable Monte Carlo result takes a substantially longer time (the solid line shows the average and the shaded area depicts the confidence interval for Monte Carlo simulations). In particular, it takes on average more than  $200ms$  to achieve a 20% error and more than  $2000ms$  are required to guarantee a 5% error (which is more than 4600 times slower than our exact algorithm).

Figures 4.4 - 4.6 concern game  $g_2$  for different values of  $k$  ( $k = 2$ ,  $k_i = \frac{deg_i}{2}$ , and  $k_i = \frac{3}{4} deg_i$ , respectively, where  $deg_i$  is the degree of node  $v_i$ ).<sup>13</sup> The advantage of our exact algorithm over Monte Carlo simulation is again exponential.

Replacing the distance threshold  $d_{cutoff}$  with a hop threshold  $k_{cutoff}$  enables game  $g_3$  to be played on an unweighted network. Thus, similarly to games  $g_1$  and  $g_2$ , we test it on the Western States Power Grid. The results are shown on Figures 4.7 and 4.8 for  $k_{cutoff}$  being equal to 2 and 3, respectively. The third game is clearly more computationally challenging than  $g_1$  and  $g_2$  (note that the vertical axis is in seconds instead of milliseconds). Now, our exact algorithm takes

<sup>13</sup> Recall that in  $g_2$  the meaning of parameter  $k$  is as follows: the value of coalition  $C$  depends on the number of nodes in the network with at least  $k$  neighbours in  $C$ .

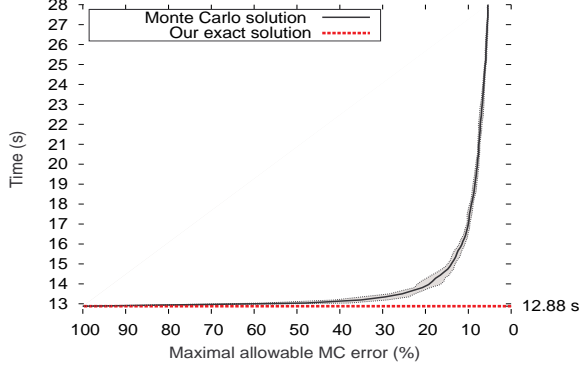


Figure 4.7:  $g_3, k_{cutoff} = 2$ , WSPG (UW)

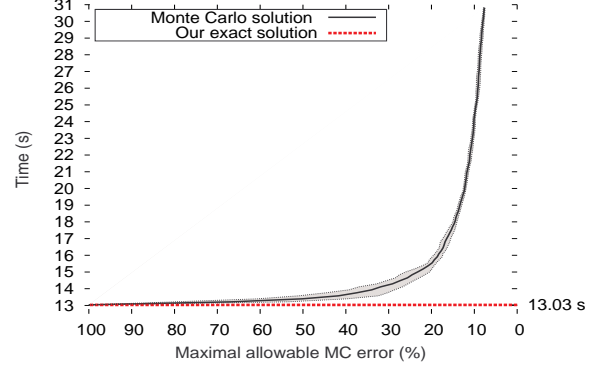


Figure 4.8:  $g_3, k_{cutoff} = 3$ , WSPG (UW)

about 13s to complete. The much lower speedups of the exact methods with respect to Monte Carlo approach stem from the fact that both algorithms have to start with Dijkstra's algorithm. Although this algorithm has to be run only once in both cases it takes more than 12.5s for the considered network. This means that the exact solution is slower by orders of magnitude (compared to games  $g_1$  and  $g_2$ ). The Monte Carlo approach is also slower, but this slowdown is much less significant in relative terms.

Figures 4.9 and 4.10 show the performance of the algorithms for game  $g_3$  on the astrophysics collaboration network that, unlike the Western States Power Grid, is a weighted network. We observe that increasing the value of  $d_{cutoff}$  (here from  $d_{cutoff} = \frac{d_{avg}}{8}$  to  $d_{cutoff} = \frac{d_{avg}}{4}$ ) significantly worsens the performance of the Monte Carlo-based algorithm. This is because the increasing number of nodes that have to be taken into account while computing marginal contributions (see the inner loop in Algorithm 8) is not only more time consuming, but also increases the Monte Carlo error.

For game  $g_4$  the performance of algorithms is shown in Figures 4.11 - 4.13 (for  $f(d) = \frac{1}{1+d}$ ,  $f(d) = \frac{1}{1+d^2}$  and  $f(d) = e^{-d}$ , respectively). Whereas the Monte Carlo methods for the first three games are able to achieve a reasonable error bound in seconds or minutes, for the fourth game it takes more than 40 hours to approach 50% error. This is because the inner loop of the *Marginal Contribution* block (see Algorithm 9) iterates over all nodes in the network. Due to the time consuming performance we run the simulations only once. Interestingly, we observe that the error of the Monte Carlo method sometimes increases slightly when more iterations are performed. This confirms that the error of the Monte Carlo method to approximate the Shapley value proposed in Castro et al. [2009] is only statistically decreasing in time. Certain new randomly chosen permutations can actually increase the error.

Figures 4.14, 4.15, 4.16 and 4.17 present comparisons of our approximation algorithm for game  $g_5$  against Monte Carlo sampling for small networks (for which the exact Shapley value can be computed from the Definition 5)). In these figures, the horizontal dotted line shows the

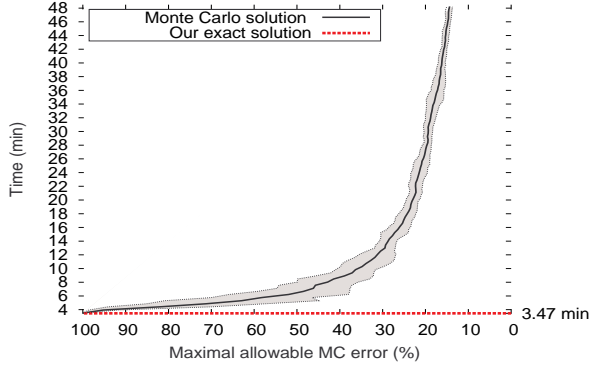


Figure 4.9:  $g_3, d_{cutoff} = \frac{d_{avg}}{8}$ , APhC (W)

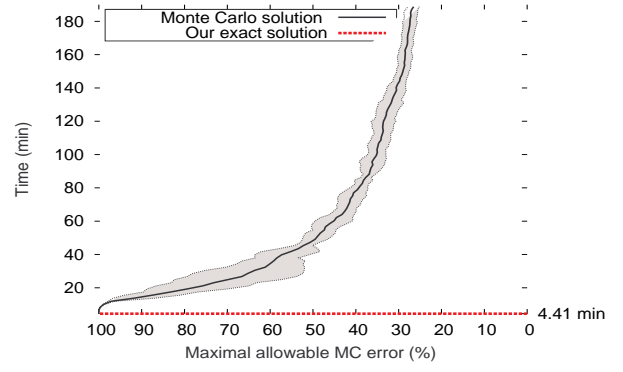


Figure 4.10:  $g_3, d_{cutoff} = \frac{d_{avg}}{4}$ , APhC (W)

running time of our solution, while the vertical dotted line shows its average approximation error with the shaded area being the confidence interval. As previously, the solid line shows the average, and the shaded area depicts the confidence interval for the Monte Carlo simulations. We see in Figures 4.14, 4.15 and 4.16 that the approximation error in our proposed algorithm is well-contained for small networks. Specifically, for  $K_6$  it is about 10%; whereas for the bigger network  $K_{12}$  it is about 5%. However, we notice that, for higher values of  $W_{cutoff}$ , the Monte Carlo method may slightly outperform our solution. See in Figure 4.15 how the average approximation error of the Monte Carlo sampling achieved in  $0.38ms$  is lower than the average error achieved by our method. Already for  $K_{12}$  this effect does not occur (see Figure 4.17).

For large networks, where the exact Shapley value cannot be obtained, we are naturally unable to compute exact approximation error. We believe that this error may be higher than the values obtained for  $K_6$  and  $K_{12}$ . However, the mixed strategy, that we discussed in Section 4.3.5 and that uses our approximation only for large degree vertices, should work towards containing the error within practical tolerance bounds. As far as we believe that Monte Carlo gives good results, from Figure 4.18, we can infer that our approximation solution for large networks gives good results (within 5%) and is at least two times faster than the Monte Carlo algorithm.

To summarise, our exact solutions outperform Monte Carlo simulations even if relatively wide error margins are allowed. However, this is not always the case for our approximation algorithm for game  $g_5$ . Furthermore, it should be underlined that if the centrality metrics under consideration cannot be described with any of the games  $g_1$  to  $g_4$  for which exact algorithms are now available, then Monte Carlo simulations are still a viable option.

## 4.5. Conclusions

The key finding of this chapter is that the Shapley value for many centrality-related cooperative games of interest played on networks can be solved analytically. In particular, we proposed

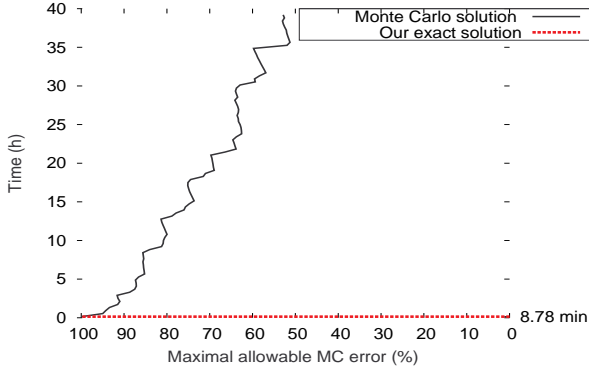


Figure 4.11:  $g_4, f(d) = \frac{1}{1+d}, \text{APhC (W)}$

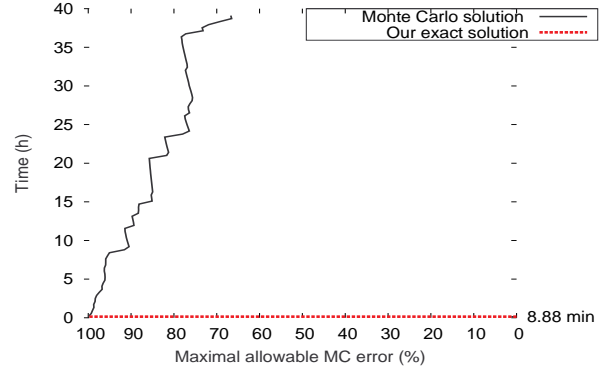


Figure 4.12:  $g_4, f(d) = \frac{1}{1+d^2}, \text{APhC (W)}$

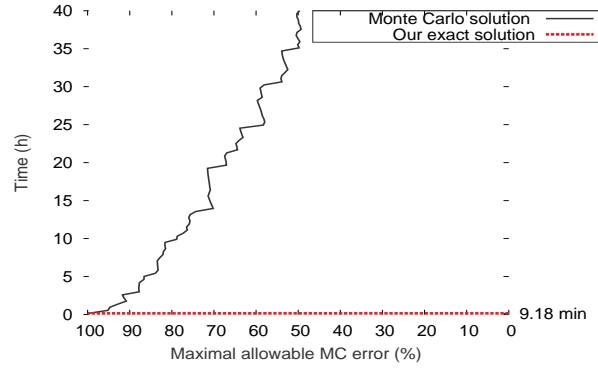


Figure 4.13:  $g_4, f(d) = e^{-d}, \text{APhC (W)}$

the polynomial time algorithms for computing Shapley value-based degree and closeness centralities. The resulting algorithms are not only error-free, but also run in polynomial time (see Table 4.1) and, in practice, are much faster than Monte Carlo methods. Approximate closed-form expressions and algorithms can also be constructed for some classes of games played on weighted networks. Simulation results show that these algorithms are acceptable for a range of situations involving big networks.

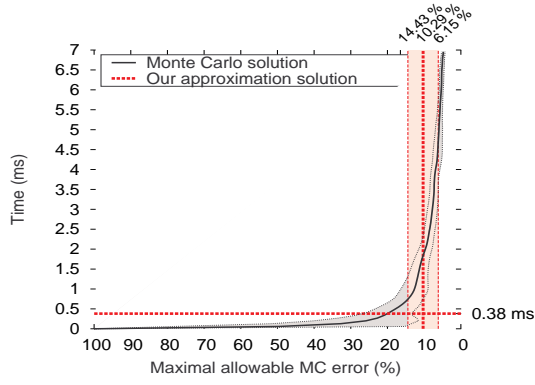


Figure 4.14:  $g_5$ ,  $W_{cutoff} = \frac{1}{4}\alpha_i$ ,  $K_6$  (W)

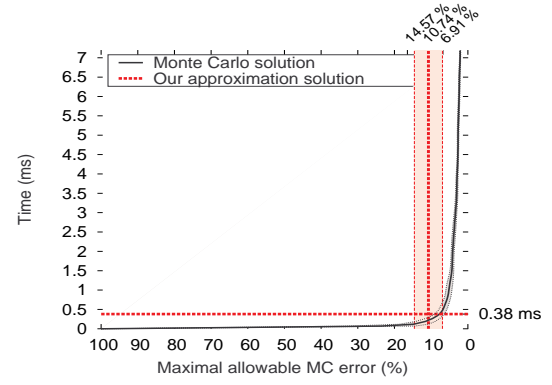


Figure 4.15:  $g_5$ ,  $W_{cutoff} = \frac{3}{4}\alpha_i$ ,  $K_6$  (W),

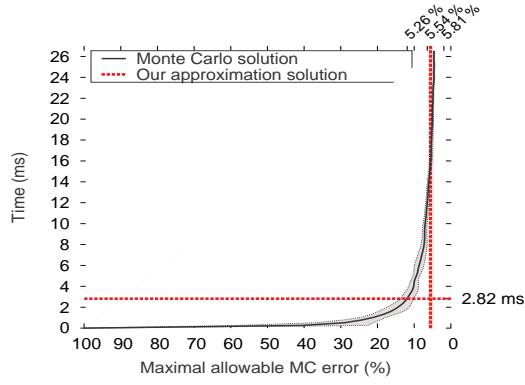


Figure 4.16:  $g_5$ ,  $W_{cutoff} = \frac{1}{4}\alpha_i$ ,  $K_{12}$  (W)

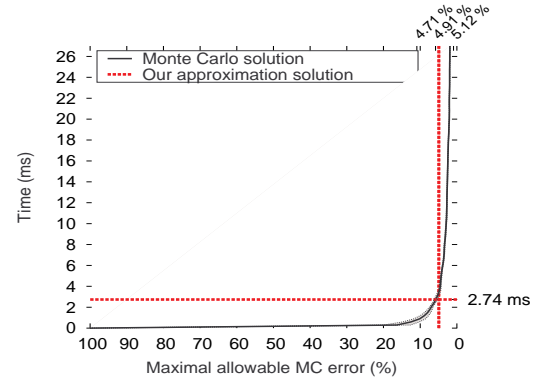


Figure 4.17:  $g_5$ ,  $W_{cutoff} = \frac{3}{4}\alpha_i$ ,  $K_{12}$  (W)

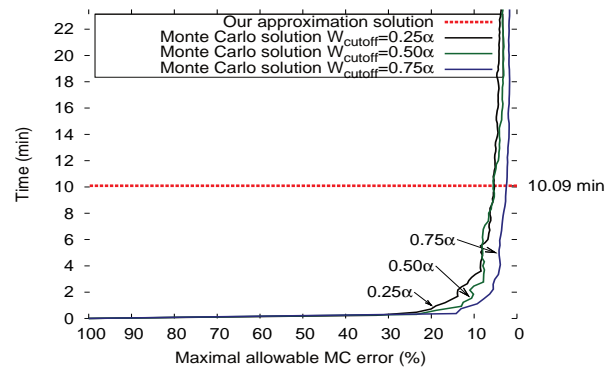


Figure 4.18:  $g_5$ ,  $W_{cutoff} = \frac{1}{4}\alpha_i$ ,  $W_{cutoff} = \frac{2}{4}\alpha_i$ , and  $W_{cutoff} = \frac{3}{4}\alpha_i$ ,  $K_{1000}$  (W)



## 5. Efficient Algorithms for a Game-Theoretic Measure of Betweenness Centrality

Betweenness centrality is a measure of the potential of individual nodes to control information flow in a network. In this chapter, we extend the standard definition of betweenness centrality and propose a family of game-theoretic measures based on *Semivalues*—a large family of solution concepts from cooperative game theory that include, among others, the Shapley value and the Banzhaf power index. The key advantage of the game-theoretic betweenness centrality over the standard measure is the ability to rank nodes while taking into account any synergies that may exist in different groups of nodes. We develop polynomial time algorithms to compute the Semivalue-based betweenness centrality in general, and the Shapley value-based betweenness centrality in particular, both in weighted and unweighted graphs. Interestingly, for the unweighted case, our algorithm for computing the Shapley value-based centrality has the same complexity as the best known algorithm due to Brandes [2001] for computing the standard betweenness centrality. The new measures are evaluated in simulated scenario where simultaneous node failures occur. Compared to the standard measure, the ranking obtained by our measures reflects more accurately the influence that different nodes have on the functionality of the network. Finally, we empirically evaluate our algorithms on weighted and unweighted random scale-free graphs, and on two real network: an email communication network, and the neural system of *C. elegans*. This evaluation shows that our algorithms are significantly faster than the fastest alternative from the literature.

All algorithms, experiments and theorems presented in this chapter are an exclusive contribution of the author of this thesis.

### 5.1. Introduction

Betweenness centrality is a measure of control introduced independently by Anthonisse [1971] and Freeman [1979], which aims to determine the importance of a given node in the context of network flow. More specifically, in its standard form, betweenness centrality considers all shortest paths that pass through the node. The more shortest paths go through the node, the more important is the role of this node in the network.

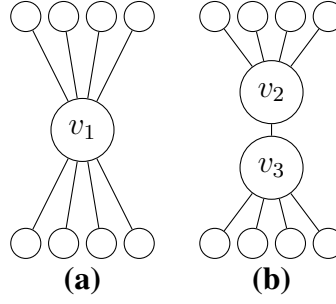


Figure 5.1: The example demonstrating that classical centrality measures do not provide any insight into synergies among the nodes. In network (a), betweenness centrality identifies  $v_1$  as the most important node. Indeed,  $v_1$  controls all 28 shortest paths between other nodes. Similarly, in network (b), nodes  $v_2$  and  $v_3$  are central as they control 26 shortest paths each. Now, let us consider  $v_2$  and  $v_3$  as a group  $\{v_2, v_3\}$ . In terms of network flow, the role of this group in network (b) is exactly the same as the role of  $v_1$  in network (a). However, it is unknown how the role of this group could be derived from the classical betweenness centrality (of both individual nodes). For instance, one natural approach to quantify the role of a group of nodes would be to sum up their individual centralities. But, for  $\{v_2, v_3\}$ , this would yield  $26 + 26 = 52 > 28$  which clearly exaggerates their joint value. Another simple approach would be to average the classical centrality  $v_2$  and  $v_3$  [Everett and Borgatti, 1999]. This would however understate the role of  $\{v_2, v_3\}$  since  $\frac{26+26}{2} = 26 < 28$ . Now, the group betweenness centrality assigns to  $\{v_2, v_3\}$  value of 28. Thus, unlike classical betweenness centrality, the group betweenness centrality is able to capture the negative synergy between  $v_2$  and  $v_3$ .

Betweenness centrality (Definition 16) is one of the classical measures of centrality in network analysis, with the other measures being degree centrality, closeness centrality, and eigenvector centrality (see Section 2.7). Several applications of betweenness centrality have been considered in the literature [Puzis et al., 2007b, Dolev et al., 2010]. It has also been implemented in numerous software tools for network analysis [Brandes, 2008]. Furthermore, a variety of extensions of betweenness centrality have been proposed in order to suit the needs of particular applications. For instance, a well-known community detection algorithm by Girvan and Newman [2002] is based on edge betweenness centrality.

One of the characteristics of betweenness centrality, as well as the other classical centrality measures, is that it focuses on the role played by each node individually. However, such an approach can be inadequate if there exist non-negligible positive or negative synergies between nodes when their role is considered in groups. Simply put, classical centrality measures are not able to account for such synergies.

To address this problem, Everett and Borgatti [1999] introduced *group centrality* (see Definition 22). While group centrality measures have the advantage of taking synergies into account, these measures have their own shortcomings. Specifically, they produce a ranking of an exponential number of groups, which is hard to maintain in memory, and hard to reason about, as it is not obvious how to extract meaningful conclusions from so many groups. What would

be more manageable is to produce a ranking of the nodes themselves, based on the ranking of all possible groups. The answer to it was given in the previous chapter where game-theoretic centrality measures was introduced (Definition 25).

A potential downside of *game-theoretic network centralities* is that many methods from cooperative game theory are computationally challenging. For instance, given a coalitional game defined over a network of  $|V|$  nodes, a straight-forward computation of the Shapley value requires considering all possible  $2^{|V|}$  coalitions (i.e., groups) of nodes. This is clearly prohibitive for networks with hundreds, or even tens, of nodes. Indeed, it has been shown that in some cases the exponential number of computations cannot be avoided, i.e., it is impossible to compute particular game-theoretic network centralities in time polynomial in the size of the network.

Fortunately, the positive computational results have been given in the previous chapter where we analysed various game-theoretic extensions of degree and closeness centrality and showed that sometimes it is possible to leverage the fact that the synergies in coalitions depend on the topology of the network. As a result, it was showed that some game-theoretic centrality measures are indeed computable in polynomial time.

Nevertheless, the aforementioned positive results concerned solely the extensions of degree and closeness centralities which are, in general, computationally less challenging than betweenness centrality. In fact, no game-theoretic extension of betweenness centrality has been developed to date.

In this chapter, we fill the above gap in the literature by proposing and analysing the computational aspects of the *game-theoretic betweenness centrality*. This is the first game-theoretic centrality that is based on *Semivalues*—a family of generalisations of the *Shapley value* that offer a wider spectrum of ways in which the role of nodes within various groups can be evaluated.

As our main contribution, we propose polynomial time algorithms to compute both the Shapley value-based and Semivalue-based betweenness centralities, for weighted graphs as well as unweighted graphs. Surprisingly, as shown in Table 5.1, for the unweighted case, our algorithm for computing the Shapley value-based centrality has the same complexity as the best known algorithm due to Brandes [2001] for computing the standard betweenness centrality, i.e.,  $O(|V||E|)$ , where  $V$  is the set of nodes and  $E$  set of edges in the network. Moreover, for Semivalues which are, in general, more computationally challenging than the Shapley value, the running time of our algorithm is  $O(|V|^2|E|)$  (for the unweighted graphs).

We consider three applications of the new centrality measures. The first one concerns the protection of computer network against Spyware—the malicious computer programs, the second example is about the issue of measuring systemic importance of financial institutions [European Central Bank, 2010]. Under certain assumptions, the Shapley value- and the

Semivalued-based centrality measures turn out to be the exact solution to this problem. In the third application we simulate simultaneous node failures in a network. Following Albert et al. [2000] and Holme et al. [2002], among others, we quantify the functionality of the network based on the average inverse geodesic measure. The simulation results show that compared to the standard measure, the cardinal ranking obtained by our measures reflects more accurately the influence that different nodes have on the network functionality.

## 5.2. The Overview of Computational Results

The literature review on centralities and game-theoretic centrality in particular, was already presented in Chapter 3. Now, we discuss the computational properties of the classical betweenness centrality. In particular, the fastest known dedicated algorithm to compute it is due to Brandes [2001]. The algorithm works in  $O(|V||E|)$  time for unweighted graphs, and in  $O(|V||E| + |V|^2 \log(|V|))$  time for weighted graphs, where  $V$  is the set of nodes and  $E$  set of edges in the network. Furthermore, group betweenness centrality for a pre-defined group of nodes can be computed in  $O(|V||E|)$  time as shown by Brandes [2008]. Puzis et al. [2007b] developed an algorithm to compute group betweenness centrality of a given subset  $S$ . The algorithm starts with a preprocessing step (which alone takes  $O(|V|^3)$  time), and then computes group betweenness centrality in  $O(|S|^3)$  time.

Centrality	Unweighted Graphs	Weighted Graphs
standard betweenness	$O( V  E )$ Brandes [2001]	$O( V  E  +  V ^2 \log( V ))$ Brandes [2001]
group betweenness	$O( V  E )$ Brandes [2008]	$O( V  E  +  V ^2 \log( V ))$ Brandes [2008]
Shapley value-based betweenness	$O( V  E )$ <b>this dissertation</b>	$O( V ^2 E  +  V ^2 \log V )$ <b>this dissertation</b>
Semivalued-based betweenness	$O( V ^2 E )$ <b>this dissertation</b>	$O( V ^3 E  +  V ^3 \log V )$ <b>this dissertation</b>

Table 5.1: Summary of the computational results obtained in this chapter.

Table 5.1 summarises the above complexity results (for betweenness centrality and group betweenness centrality) as well as our computational results (for Shapley-based and Semivalued-based betweenness centralities). As can be seen, our algorithms also run in polynomial time, both for unweighted and weighted graphs. Observe that our algorithm for computing the Shapley value-based betweenness centrality for unweighted graphs has the same complexity as the aforementioned algorithm by Brandes [2001] for standard betweenness centrality, which is surprising given the much more complicated nature of our measure.

### 5.3. The New Centrality and Its Properties

In this section we introduce a family of game-theoretic betweenness centrality measures and discuss their basic properties. We start in Section 5.3.1 with a few motivating scenarios that demonstrate the need for a more involved centrality measure than standard betweenness centrality. Sections 5.3.2 and 5.3.3 introduce our Shapley value-based and Semivalue-based betweenness centralities, respectively. Finally, sections 5.3.4 and 5.3.5 analyse the marginal contributions of nodes, in the context of the Shapley value-based and Semivalue-based betweenness centralities, respectively.

#### 5.3.1. Motivating Scenarios

We will consider three motivating scenarios: protect computer network against malicious programs (Section 5.3.1), ranking institutions in a financial system according to their systemic importance (Section 5.3.1), and analysing network vulnerability (Section 5.3.1).

#### Detecting Intrusion in a Network

One of the scenarios on detecting intrusion in a network considered in the literature is *distributed network intrusion detection* [Valdes and Skinner, 2001]. In a nutshell, the aim here is to filter the content of a network by installing some anti-spyware software on a chosen set of  $k$  nodes. Now, assuming that the majority of the network flow is transmitted through shortest paths, and ignoring the capacities of cables and routers, the best choice is to install the anti-spyware on the set of nodes  $S \subseteq V, |S| = k$  whose group betweenness centrality  $c_{gb}(S)$  (Definition 22) is maximal. Unfortunately, the problem is NP-hard; nonetheless, there exists approximation algorithms to solve it [Puzis et al., 2007a].

The solution to the above problem can be used to guide the efforts to protect the most important nodes against a potential spyware attack. However, how to choose the optimal points to protect if we have limited resources and do not know in advance the size of the attack? To model such a situation, let us consider the following scenario. Suppose that nodes in a communication network have been exposed to the possibility of a random spyware infection. The more shortest paths there are that pass through the infected group, the worse it is for the network, because the malicious software may control or monitor the flow. Furthermore, suppose that neither the subset of infected nodes is known nor the size of the attack (i.e., the number of infected nodes). Assuming that the administrators' resources are limited and that the malicious software can be located at any workstation, a decision has to be made on which node to examine first. In other words, we need to identify a node that makes the biggest contribution to the spyware's ability to monitor the flow going through the infected nodes. Hence, we formalize this problem as follows:

**Problem 1.** Given a discrete probability distribution  $D : \{1, \dots, n\} \rightarrow [0, 1]$  and a random subset  $S : |S| \sim D$ , which node  $v \in V$  in the graph  $G = (V, E)$  maximizes the expression:  $c_{gb}(S) - c_{gb}(S \setminus \{v\})$ .

Now if it was known that exactly one node has been affected by the spyware attack ( $|S| = 1$ ) and that the single target was chosen uniformly at random, then the standard betweenness centrality would solve the above problem; the node with the highest such centrality is the solution. However, if there is a possibility that multiple nodes have been infected ( $|S| > 1$ ), then the standard betweenness centrality becomes insufficient. This is because betweenness centrality of a group of nodes is not necessarily equal to the sum of betweenness centralities of individual nodes within this group (see Figure 5.2 for an example). In other words, the following does not necessarily hold for every  $S \subseteq V$ :

$$c_{gb}(S) = \sum_{v \in V} c_b(v).$$

Hence, a node  $v \in V$  that maximizes the expression  $c_{gb}(S) - c_{gb}(S \setminus \{v\})$  for some  $S \subseteq V$  does not have to be the one with the highest standard betweenness centrality.

In this chapter we propose the polynomial-time exact solution to Problem 1 for a family of probability distributions of choosing the set  $S$ .

## Measuring Systemic Importance of Financial Institutions

The crisis of year 2007 and 2008 clearly showed the need for more stringent supervision of financial systems. As a result, a number of systemic risk monitoring and supervisory bodies have been created worldwide: the Financial Stability Oversight Council (USA), the European Systemic Risk Board and the Financial Stability Board. Furthermore, new prudential policies and regulations were laid out. Most notably, the Basel III accords aim at introducing micro- and macro-prudential regulations such as stricter capital reserve ratios, increased audit transparency and more thorough risk management by banks and other financial institutions.

At a discretion of supervising authorities, systemically important institutions may be asked to follow additional macro-prudential policies including a combination of capital surcharges, contingent capital and bail-in debt. As stated in the Financial Stability Board's interim report to G20 leaders, "*Financial institutions should be subject to requirements commensurate with the risks they pose to the financial system*" [Financial Stability Board, 2010].

In this context, betweenness is considered to be the most suitable centrality to measure systemic importance of institutions in the financial system [European Central Bank, 2010]. In more detail, it has the following natural interpretation: "*the betweenness of a bank A connecting pairs of nodes in the network is a measure of the dependence of these other banks from A to transfer the loans.*" [Gabrieli, 2011]. As a result, the higher betweenness centrality of an

institution is, the more attention and scrutiny it requires. It is feared that in the worst case, even one of the institution with the highest betweenness centrality may be able to bring down the entire system, due to contagion of illiquidity [Freixas et al., 2000].

However, imposing additional capital requirements on financial institutions according to their standard betweenness centrality ignores the fact that financial crises are often ignited not by factors specific to a particular bank but by (combinations of) factors that simultaneously affect possibly many banks in the system. The infamous financial crises of 2009 started with the collapse of the subprime residential mortgage market in the United States and then spread to the rest of the world, due to the exposure to American real estate assets via a complex array of financial derivatives. As a result, many financial institutions in various countries asked for urgent state interventions [Laeven and Valencia, 2010].

Consequently, it would be desirable to extend betweenness centrality so that it quantifies the betweenness of an individual institution but taking into account its role in various groups of institutions, as they can be affected simultaneously. We formalize this problem as follows:

**Problem 2.** *Given graph  $G = (V, E)$ , what is the (cardinal) ranking of nodes that reflects the marginal contribution of an individual node to group betweenness centrality of a random subset of nodes  $S : |S| \sim D$ , i.e.,  $c_{gb}(S) - c_{gb}(S \setminus \{v\})$ , where  $D : \{1, \dots, n\} \rightarrow [0, 1]$  is a discrete probability distribution.*

This problem extends the Problem 1 in the way, that we consider the whole cardinal ranking of nodes, instead of looking only for the most important individual. In this chapter, we propose the polynomial-time exact solution to Problem 2 for a family of probability distributions  $D$  of choosing set  $S$ .

### Analysis of Network Vulnerability

One of the extensive research directions on betweenness centrality involves the analysis of network vulnerability [Bompard et al., 2010, Holme et al., 2002]. The aim is to identify (and possibly protect) the most critical nodes, whose removal degrades the functionality of the network the most. In this context, a standard measure used to assess the condition of the network is called *the average inverse geodesic measure* which, essentially, is a sum of inverse distances between any pair of nodes in the network [Holme et al., 2002, Albert et al., 2000]. Formally:

$$\text{IGM}(V, E) = \sum_{v \in V} \sum_{u \in V, u \neq v} \frac{1}{d(v, u)}$$

where  $d(u, v)$  is the distance, i.e., the length of the shortest path, between nodes  $u$  and  $v$  [Freeman, 1979].

In order to formalize the problem of network vulnerability, let us first note that it is possible to reformulate Problem 1 by assuming that we are looking for a set of nodes  $C \subseteq V$  of a

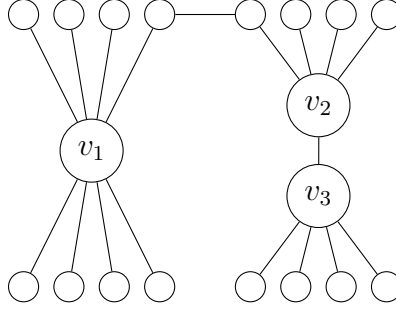


Figure 5.2: In the above sample network, both  $v_1$  and  $v_2$  are ranked equally by standard betweenness centrality;  $c_b(v_1) = c_b(v_2) = 98$ . Indeed, either  $v_1$  or  $v_2$  control the same number of shortest paths in this network. However, if we consider the role of nodes played within various groups—as is the case in our scenario where financial difficulties may affect multiple institutions simultaneously—then  $v_1$  and  $v_2$  should not be evaluated equally. To see why this is the case, let us consider nodes  $v_2$  and  $v_3$ . A significant percentage of the shortest paths controlled by  $v_2$  are also controlled by  $v_3$ . Consequently, the ability to control financial flows by the group of banks  $\{v_2, v_3\}$  is smaller than by the group  $\{v_1, v_3\}$ . As a result,  $v_1$  is more systemically important than  $v_2$ .

pre-defined size  $|C| = c$ . This set should maximize the expression:  $c_{gb}(S) - c_{gb}(S \setminus C)$ , where  $S$  is a random subset of nodes  $S \subseteq V$  such that  $l \leq |S| \leq k$ .

Now, we define the problem of network vulnerability as follows:

**Problem 3.** *Given a discrete probability distribution  $D : \{1, \dots, n\} \rightarrow [0, 1]$  and a random subset  $S : |S| \sim D$ , which set of nodes  $C \subseteq V$  of a given size  $|C| = c$  in the graph  $G = (V, E)$  maximizes the expression:  $IGM(V \setminus S) - IGM(V \setminus (S \setminus C))$ .*

where  $IGM$  stands for the average inverse geodesic measure.

To date, standard betweenness centrality has been advocated as an effective approximation that solves Problem 3 [Holme et al., 2002]. However, in Section 5.5, we experimentally show that the game-theoretic betweenness centrality achieves consistently better results than standard betweenness centrality.

### 5.3.2. The Shapley Value-based Betweenness Centrality

In this subsection we will define group betweenness centrality. Firstly, recall that in our approach, we cast a network into the combinatorial structure of a coalitional game. That is, we consider all subsets of the set of nodes  $V(G)$  and we define a function that assigns to every such subset a numerical value; this value is in fact the subset's group betweenness centrality. Formally, borrowing notation from coalitional games, we define a tuple  $(V(G), \nu)$ , where function  $\nu : 2^{V(G)} \rightarrow \mathbb{R}$  assigns to each  $S \subseteq V(G)$  the value  $c_{gb}(S)$ , and  $\nu(\emptyset) = 0$ . This means that  $\nu$  returns the group betweenness centrality of any given subset of nodes. Now since  $(V(G), \nu)$  has the same combinatorial structure as that of a coalitional game, we can use the Shapley



value to quantify the importance of nodes based on the impact their membership makes on group betweenness centrality. In other words, we obtain a measure of centrality of individual nodes which aggregates information about their role in group betweenness centrality across various groups. Formally, the Shapley value-based betweenness centrality measure is defined as follows:

**Definition 28 (Shapley value-based betweenness centrality measure).** *Shapley-value based betweenness centrality is a pair  $(\psi_B, \phi^{SV})$  consisting of a representation function  $\psi_B(G) = c_{gb}$  that maps each graph to the game with group betweenness centrality and a Shapley value solution concept  $\phi^{SV}$ .*

**Example 6.** Consider again nodes  $v_1$  and  $v_2$  from Figure 5.2. Here:

$$\phi_1^{SV}(V(G), \psi_B(G)) = 18.2 \quad \text{and} \quad \phi_2^{SV}(V(G), \psi_B(G)) = 16.0833.$$

Unlike the standard betweenness centrality, the ranking produced by our Shapley-value based betweenness centrality reflects the fact that node  $v_2$  has many common shortest paths with  $v_3$ , while  $v_1$  does not.

Intuitively, the Shapley value-based betweenness centrality measure represents the load placed on a given node, taking into account the role it plays in all possible groups of nodes in the network. Moreover, the Shapley value-based betweenness centrality satisfies the same properties as the Shapley value (see Section 2.2). Let us translate them to the network context:

1. **Symmetry**—any two symmetric nodes  $v_i$  and  $v_j$  are assigned the same centrality, where symmetry refers here to the fact that the two nodes can be interchanged in any group of nodes without affecting the group's betweenness. Formally, if  $v(S \cup \{v_i\}) = v(S \cup \{v_j\}) \forall S \subseteq V \setminus \{v_i, v_j\}$ , then:  $\phi_i^{SV} = \phi_j^{SV}$ ;
2. **Marginality**—if a node's contribution to the group betweenness centrality of each subset of nodes in network  $G = (V, E)$  is no less than its contribution to the same subset in another network  $G' = (V, E')$ , then its centrality in the former network should be no less than in the latter one. Formally, for every pair of networks,  $G = (V, E)$  and  $G' = (V, E')$ , and functions  $\nu_G$  and  $\nu_{G'}$ , defined as in equation (22), and for every node  $v_i \in V$ , if

$$\nu_G(S \cup \{v_i\}) - \nu_G(S) \geq \nu_{G'}(S \cup \{v_i\}) - \nu_{G'}(S), \quad \forall S \subseteq V \setminus \{v_i\},$$

then  $\phi_v^{SV}$  in  $G$  is not smaller than  $\phi_v^{SV}$  in  $G'$ .

3. **Efficiency**—this property simply says that the sum of the Shapley value-based betweenness centralities for all the nodes in the network is normalised to 0, i.e.,  $\sum_{v_i \in V} \phi_i^{SV} = 0$ .<sup>1</sup>

---

<sup>1</sup> Note that the group betweenness centrality of the entire network, i.e., of the group  $S = V$ , equals 0.

Naturally, it is also possible to normalize the Shapley value-based betweenness centrality so it ranges from 0 to 1. To this end, one should use  $\frac{\phi_i^{SV}}{2 \max_{v_i \in V} c_b(v_i)} + \frac{1}{2}$ . This is because  $\forall_{v_i \in V} \forall_{C \subseteq V} (c_{gb}(C \cup \{v_i\}) - c_{gb}(C)) \leq c_b(v_i)$ .

Observe that the new centrality solves Problem 1 and Problem 2 exactly for a particular probability distribution  $PD$ .

**Proposition 3.** *The Shapley value-based betweenness centrality is the exact solution to Problem 1 and Problem 2 in a particular case in which each set  $S \subseteq V$  of size  $|S| = k > 0$  is chosen with probability  $\frac{1}{|V| \binom{|V|-1}{k-1}}$ .*

*Proof.* To show it, it is enough to rewrite formula (2.3) as follows:

$$\phi_i^{SV}(V, \nu) = \frac{1}{|V|} \sum_{1 \leq k \leq |V|} \frac{1}{\binom{|V|-1}{k-1}} \sum_{\substack{S \subseteq V \\ |S|=k \\ v_i \in S}} (\nu(S) - \nu(S \setminus \{v_i\})) \quad (5.1)$$

The inner sum in the above equation runs through all subsets of nodes containing  $v_i$  of a given size  $k$ , and  $\binom{|V|-1}{k-1}$  is the number of such subsets. The first fraction implies that the size  $k \in \{1, \dots, n\}$  is chosen uniformly at random. The second fraction implies that out of all subsets of size  $k$ , one is chosen uniformly at random. ■

Formula (5.1) has the following interpretation in the context of measuring systemic importance of financial institutions:

1. Fraction  $\frac{1}{|V|}$  means that the probability that the financial crises starts with a group of financial institutions of size  $|S| = k$  is the same for any  $k$  (where  $k \in \{1, \dots, |V|\}$ ). We will denote this probability  $PD(V, k)$ .
2. Fraction  $\frac{1}{\binom{|V|-1}{k-1}}$  means that the probability that each particular  $k$ -sized subset of institutions containing  $v_i$  bankrupts is the same.

While the latter implication seems natural, the former implication seems to be a special case. Arguably, it does not seem natural to assume that the probability of a single institution bankruptcy is exactly the same as the probability of having all institutions bankrupted (especially if the network is very large). To relax this assumption, we need to change the fraction  $\frac{1}{|V|}$  in equation (5.1). However, this fraction is inherent in the Shapley value. As such, modifying it requires adopting Semivalues as a centrality measure. In the next section, we show how this can be done.

### 5.3.3. Semivalue-based Betweenness Centrality

Recall that a Semivalues defined in equation (2.5) generalises the Shapley value by allowing for different sequences of weights  $p_{|S|}$  for each set  $S$ . Building upon Semivalues, we will now introduce a centrality measure that generalizes the Shapley value-based betweenness centrality.

**Definition 29 (Semivalue-based betweenness centrality measure).** *Semivalue-based betweenness centrality measure is a pair  $(\psi_B, \phi^{SEMI})$  consisting of a representation function  $\psi_B(G) = c_{gb}$  that maps each graph to the game with group betweenness centrality and a Semivalue  $\phi^{SEMI}$ .*

We observe that the above family of Semivalue-based betweenness centrality measures inherits from Semivalues the properties of symmetry and marginality. Naturally, particular centralities may satisfy additional properties inherited from the corresponding solution concepts in coalitional games. For instance, it is not difficult to see that, similarly to the Banzhaf index [Casajus, 2011], the Banzhaf index-based betweenness centrality measure satisfies the “2-Efficiency” property. In essence, this property says that, for each pair of nodes  $u, v \in V$ , if you consider them to be a single node  $vu$  (note that  $v$  and  $u$  are not merged, but they appear in each coalition together) then the Banzhaf index-based betweenness centrality of  $vu$  equals the sum of the Banzhaf index-based betweenness centralities of  $u$  and  $v$ .

**Proposition 4.** *The Semivalue-based betweenness centrality is the exact solution to Problem 1 and Problem 2, where each set  $S \subseteq V$  of size  $|S| = k > 0$  is chosen with probability  $PD(V, k) \frac{1}{\binom{|V|}{k}}$ .*

*Proof.* To observe it, we transform formula (2.5) as follows:

$$\phi^{SEMI}(V, \nu) = \sum_{1 \leq k \leq |V|} PD(V, k) \frac{1}{\binom{|V|-1}{k-1}} \sum_{\substack{S \subseteq V \\ |S|=k \\ v_i \in S}} (v(S) - v(S \setminus \{v_i\})), \quad (5.2)$$

where  $PD(V, k) = p_{k-1} \binom{|V|-1}{k-1}$ .

The analysis of the above equation leads to the following. Each set  $v_i \in S \subseteq (V \setminus \{v_i\})$  such that  $|S| = k$  is chosen with probability  $PD(V, k)$ , and each group of equal size is chosen with equal probability, i.e.,  $PD(V, k) \cdot \frac{1}{\binom{|V|-1}{k-1}}$ . ■

In our motivating scenario of measuring systemic importance of financial institutions, with the Semivalue-based betweenness centrality we are able to assume an arbitrary probability distribution on the size of the bankrupting institution.

Also the Semivalue-based between centrality can be normalized to the interval  $[0, 1]$  in the same manner as Shapley value-based betweenness centrality. To this end, one needs to compute:

$$\frac{c_S(v_i)}{2 \max_{v_i \in V} c_b(v_i)} + \frac{1}{2}$$

In the following subsections, we provide formal analysis of the Shapley value-based betweenness centrality and the more-general Semivalue-based betweenness centrality.

#### 5.3.4. A Look at Marginal Contribution for the Shapley Value

The centrality of a node is measured by computing a weighted average of its marginal contribution to the group-betweenness centrality of all subsets of nodes in the graphs. In this subsection, we show how to compute this weighted average efficiently.

To this end, we will use Definition 6. More specifically, given a graph  $G$  and some vertex  $v \in V(G)$ , we would like to compute the expected marginal contribution of this vertex to the set of vertices  $P_\pi(v)$  that precede  $v$  in a random permutation of all vertices of the graph. We focus in our analysis on three exhaustive cases, where the marginal contribution is positive, negative, or neutral, respectively.

First, let us consider positive contributions, and let us focus on some particular shortest path  $p$  which contains vertex  $v$ . Recall that we denote by  $\sigma_{st}$  the number of shortest paths between vertices  $s$  and  $t$ . Furthermore, we denote by  $\sigma_{st}(v)$  the number of shortest paths between vertices  $s$  and  $t$  that each pass through vertex  $v$ , where  $t \neq v \neq s$  (recall that every such path is said to be *controlled* by  $v$ ). Every path in  $\sigma_{st}(v)$  has a positive contribution to the coalition  $P_\pi(v)$  via vertex  $v$  if and only if the path is not controlled by (i.e., does not pass through) any vertex from the set  $P_\pi(v)$ . In this case, the positive contribution equals  $\frac{1}{\sigma_{st}}$ . The necessary and sufficient condition for this to happen can be expressed by  $\Psi(p) \cap P_\pi(v) = \emptyset$ , where  $\Psi(p)$  is the set of all vertices lying on the path  $p$ , including endpoints.

Now, let us introduce a Bernoulli random variable  $B_{v,p}^+$  which indicates whether vertex  $v$  makes a positive contribution to the set  $P_\pi(v)$  through shortest path  $p$ . Thus, we have:

$$\mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] = \frac{1}{\sigma_{st}} P[\Psi(p) \cap P_\pi(v) = \emptyset],$$

where  $P[\cdot]$  denotes probability, and  $\mathbb{E}[\cdot]$  denotes expected value. In other words, we need to know the probability of having  $v$  precede all vertices from  $\Psi(p) \setminus \{v\}$  in a random permutation of all vertices in the graph.

**Theorem 2.** *Let  $K$  be a set of elements such that  $|K| = k$ . Furthermore, let  $L$  and  $R$  be two disjoint subsets of  $K$ , and let  $|L| = l$  and  $|R| = r$ . Now, given some element  $x \in K$ , where*

$x \notin L \cup R$ , and given a random permutation  $\pi \in \Pi(K)$ , the probability of having every element in  $L$  before  $x$ , and every element in  $R$  after  $x$ , is:

$$P[\forall_{e \in L} \pi(e) < \pi(x) \wedge \forall_{e \in R} \pi(e) > \pi(x)] = \frac{1}{(l+1) \binom{l+r+1}{r}}$$

*Proof.* Let us first count the number of ways in which a permutation  $\pi \in \Pi(K)$  can be constructed such that the following condition holds:  $\forall_{e \in L} \pi(e) < \pi(x) \wedge \forall_{e \in R} \pi(e) > \pi(x)$ . Specifically:<sup>2</sup>

- Let us choose  $l + r + 1$  positions in the permutation. There are  $\binom{k}{l+r+1}$  such choices.
- Now, in the first  $l$  chosen positions, place all the elements of  $L$ . Directly after those, place the element  $x$ . Finally, in the remaining  $r$  chosen positions, place all the elements of  $R$ . The number of such line-ups is:  $l!r!$ .
- As for the remaining elements (i.e., the elements of  $K \setminus (L \cup \{x\} \cup R)$ ), they can be arranged in the permutation in  $(k - (l + r + 1))!$  different ways.

Thus, the number of permutations satisfying our condition is:

$$\binom{k}{l+r+1} l!r! (k - (l + r + 1))! = \frac{k!}{(l+1) \binom{l+r+1}{r}}.$$

Finally, since we have a total of  $k!$  permutations, the probability that one of them satisfies our condition is given by equation (5.3). ■

Based on Theorem 2, we can obtain the probability of an event in which the vertex  $v$  lies on the path  $p$  and precedes all the other vertices from this path in a random permutation of all vertices in the graph  $G$ . Now, by setting  $K = V(G)$ ,  $L = \emptyset$  and  $R = \Psi(p) \setminus \{v\}$ , we obtain the desired probability:  $\frac{1}{|\Psi(p)|}$ . Thus:

$$\mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] = \frac{1}{\sigma_{st} |\Psi(p)|}. \quad (5.3)$$

Having discussed the first case, where marginal contributions are positive, let us now discuss the second case, where the marginal contribution of vertex  $v$  to set  $P_\pi(v)$  is negative. This happens when path  $p$  ends or starts with  $v$  (w.l.o.g. we consider  $v$  as the end point). Specifically, if coalition  $P_\pi(v)$  already controls path  $p$  (which includes vertex  $v$ ), then not only is there no value added from  $v$  becoming a member of this coalition, but there is a negative effect of this move. In particular, the group betweenness centrality assumes that a set of vertices  $S$  controls only those paths with both ends not belonging to  $S$ . Therefore, when  $v$  becomes a member of

---

<sup>2</sup> We thank the anonymous reviewer for suggesting this formulation of the proof.

coalition  $P_\pi(v)$ , its negative contribution through path  $p$  equals  $-\frac{1}{\sigma_{sv}}$ , where  $\sigma_{sv}$  denotes the number of paths that start with  $s$  and end with  $v$ .

Next, we analyse the probability of such a negative contribution of vertex  $v$ . Observe that if the marginal contribution is negative, then  $p$  must end with  $v$ . However, the opposite does not necessarily hold, i.e., if  $p$  ends with  $v$  then the marginal contribution of  $v$  is not necessarily negative; it could also be neutral. Based on this, in order to compute the probability of a negative contribution, we will start by focusing on the complementary event where  $v$  makes a neutral contribution to the set  $P_\pi(v)$ , as this simplifies the computational analysis. This complementary event happens if and only if either vertex  $s$ —the starting point of path  $p$ —belongs to set  $P_\pi(v)$ , or the path  $p$  is not controlled by any of the vertices in  $P_\pi(v)$ . Formally:

$$s \in P_\pi(v) \vee (\Psi(p) \cap P_\pi(v)) = \emptyset.$$

Based on this, we can introduce a Bernoulli random variable  $B_{v,p}^-$  indicating whether vertex  $v$  makes a negative contribution through path  $p$  to set  $P_\pi(v)$  as follows:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}} B_{v,p}^-\right] = -\frac{1}{\sigma_{sv}} (1 - P[s \in P_\pi(v) \vee (\Psi(p) \cap P_\pi(v)) = \emptyset]).$$

Again, one can show with the combinatorial arguments presented in Theorem 2 that this probability is  $P[\Psi(p) \cap P_\pi(v) = \emptyset] = \frac{1}{|\Psi(p)|}$ . Furthermore, due to symmetry, we have:  $P[s \in P_\pi(v)] = \frac{1}{2}$ . Finally, from the disjointness of the aforementioned complementary events, we have:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}} B_{v,p}^-\right] = \frac{2 - |\Psi(p)|}{2\sigma_{sv}|\Psi(p)|}. \quad (5.4)$$

Recall that our analysis was divided into three exhaustive cases, where the marginal contribution is positive, negative, or neutral. Having dealt with the first two cases, it remains to deal with the case of neutral marginal contribution. Since this case does not affect the expected marginal contribution of  $v$  to  $P_\pi(v)$ , we simply disregard it during the computation.

Next, let us compute the expected marginal contribution, by aggregating the cases of positive and negative contributions. To this end, let  $\partial_{st}$  be the set of all shortest paths from  $s$  to  $t$ , and analogously let  $\partial_{st}(v) \subseteq \partial_{st}$  be the set of shortest paths from  $s$  to  $t$  that pass through vertex  $v$ .<sup>3</sup> Now, using the expected value of the Bernoulli random variables (5.3) and (5.4) we are able to compute the Shapley value of vertex  $v$ , which is the expected marginal contribution of  $v$  to  $P_\pi(v)$ , as:

---

<sup>3</sup> Note that  $\sigma_{st} = |\partial_{st}|$  and  $\sigma_{st}(v) = |\partial_{st}(v)|$ .

$$\begin{aligned}
\phi_v^{SV}(V(G), \nu) &= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \mathbb{E}\left[\frac{1}{\sigma_{st}} B_{v,p}^+\right] + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \mathbb{E}\left[-\frac{1}{\sigma_{st}} B_{v,p}^-\right] \\
&= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \frac{1}{\sigma_{st} |\Psi(p)|} + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \frac{2 - |\Psi(p)|}{2\sigma_{sv} |\Psi(p)|}.
\end{aligned} \tag{5.5}$$

Interestingly, the above equation provides insight into the Shapley value-based betweenness centrality, which is a combination of two factors. Firstly, the left sum resembles the standard betweenness centrality, but scaled by the number of vertices that belong to each shortest path. Secondly, the right sum resembles the closeness centrality,<sup>4</sup> but with distances measured as the number of vertices on the shortest paths.

### 5.3.5. A Look at Marginal Contribution for the Semivalue

We start our analysis by rewriting equation (5.2) as:

$$\phi_v^{SEMI}(V, \nu) = \sum_{1 \leq k \leq |V|} PD(V, k) \mathbb{E}_{k-1,v}[\cdot], \tag{5.6}$$

where  $\mathbb{E}_{k-1,v}[\cdot]$  is the expected marginal contribution of vertex  $v$  to a random coalition of size  $k - 1$ . So, analogously to the previous subsection, we need to analytically compute the probability that the node  $v$  contributes to a coalition through a shortest path  $p$ . Again, here we only focus on the case of positive contribution and the case of negative contribution, since the neutral case has no impact on the expected marginal contribution. We will denote by  $S_{k-1} \subseteq V$  a random coalition of size  $k - 1$  uniformly drawn from the set  $\{S \subseteq V \setminus \{v\} : |S| = k - 1\}$ .

Firstly, we consider positive contributions. In what follows, let us focus on some particular shortest path  $p$  which contains vertex  $v$ . This path has a positive contribution to the coalition  $S_{k-1}$  through  $v$  if and only if it is not controlled by (i.e., does not pass through) any vertex from set  $S_{k-1}$ . In this case, the positive contribution equals  $\frac{1}{\sigma_{st}}$ . The necessary and sufficient condition for this to happen can be expressed by  $\Psi(p) \cap S_{k-1} = \emptyset$ , where  $\Psi(p)$  is the set of all vertices lying on the path  $p$  including endpoints.

Now, let us introduce a Bernoulli random variable  $BS_{k-1,v,p}^+$  which indicates whether vertex  $v$  makes a positive contribution through path  $p$  to set  $S_{k-1}$ . Note that, if  $|V| - |\Psi(p)| < k - 1$  then the probability of this event equals 0. Otherwise, we have:

<sup>4</sup> Recall that closeness centrality assigns to each vertex  $v$  the value of  $\sum_u \frac{1}{d(u,v)}$ .

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sigma_{st}}BS_{k-1,v,p}^+\right] &= \frac{1}{\sigma_{st}}P[\Psi(p) \cap S_{k-1} = \emptyset] = \frac{1}{\sigma_{st}} \frac{\binom{|V|-\Psi(p)}{k-1}}{\binom{|V|-1}{k-1}} \\
&= \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{st}(|V| - \Psi(p) - k + 1)!(|V| - 1)!}, \tag{5.7}
\end{aligned}$$

where  $P[\cdot]$  denotes probability, and  $\mathbb{E}[\cdot]$  denotes expected value. In other words, we need to know the probability of having random set  $S_{k-1}$  containing  $v$  and not containing any vertex from  $\Psi(p) \setminus \{v\}$ .

Having analysed the case of positive contribution, we now move on to the case of negative contribution. This happens when the path  $p$  ends with  $v$  and is controlled by the set  $S_{k-1}$ , i.e.,  $p$  passes through at least one of the nodes in  $S_{k-1}$ , where  $v \notin S_{k-1}$ . As in the previous subsection, we will simplify the analysis by focusing on the complementary event in which the marginal contribution is neutral. Observe that  $v$  makes a neutral contribution to set  $S_{k-1}$  through a path  $p \in \partial_{sv}$  if and only if either vertex  $s$ —the end point of path  $p$ —belongs to the set  $S_{k-1}$ , or the path is not controlled by any of the vertices in  $S_{k-1}$ . Formally:

$$s \in S_{k-1} \vee (\Psi(p) \cap S_{k-1}) = \emptyset.$$

Based on this, we can introduce a Bernoulli random variable  $BS_{k-1,v,p}^-$  indicating whether that vertex  $v$  makes a negative contribution through path  $p$  to set  $S_{k-1}$ . Now if  $|V| - \Psi(p) \geq k - 1$ , we obtain the following expression:

$$\begin{aligned}
\mathbb{E}\left[-\frac{1}{\sigma_{sv}}BS_{k-1,v,p}^-\right] &= -\frac{1}{\sigma_{sv}}(1 - P[s \in S_{k-1} \vee (\Psi(p) \cap S_{k-1}) = \emptyset]) \\
&= -\frac{1}{\sigma_{sv}}\left(1 - \left(\frac{k-1}{|V|-1} + \frac{(|V| - \Psi(p))!(|V| - k)!}{(|V| - \Psi(p) - k + 1)!(|V| - 1)!}\right)\right) \\
&= \frac{k-1}{\sigma_{sv}(|V|-1)} + \frac{(|V| - \Psi(p))!(|V| - k)!}{\sigma_{sv}(|V| - \Psi(p) - k + 1)!(|V| - 1)!} - \frac{1}{\sigma_{sv}}. \tag{5.8}
\end{aligned}$$

On the other hand, if  $|V| - \Psi(p) < k - 1$ , we obtain the following expression:

$$\mathbb{E}\left[-\frac{1}{\sigma_{sv}}BS_{k-1,v,p}^-\right] = \frac{k-1}{\sigma_{sv}(|V|-1)} - \frac{1}{\sigma_{sv}}. \tag{5.9}$$

Now, let us compute the expected marginal contribution, by aggregating the cases of positive and negative contributions. To this end, recall that  $\partial_{st}$  is the set of all shortest paths from  $s$  to  $t$ , and, analogously,  $\partial_{st}(v) \subseteq \partial_{st}$  is the set of shortest paths from  $s$  to  $t$  passing through vertex  $v$ . Now, using the expected value of Bernoulli random variables (5.7), (5.8) and (5.9) we are able to compute the Semivalue of vertex  $v$  by using equation (5.6):



$$\begin{aligned}
\phi_v^{SEMI} &= \sum_{0 \leq k \leq |V|-1} PD(V, k+1) \left( \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \mathbb{E} \left[ \frac{1}{\sigma_{st}} BS_{k,v,p}^+ \right] + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \mathbb{E} \left[ -\frac{1}{\sigma_{st}} BS_{k,v,p}^- \right] \right) \\
&= \sum_{1 \leq k \leq |V|} PD(V, k) \left( \sum_{s \neq v \neq t} \sum_{\substack{p \in \partial_{st}(v) \\ |V| - \Psi(p) \geq k-1}} \frac{(|V| - \Psi(p))! (|V| - k)!}{\sigma_{st} (|V| - \Psi(p) - k + 1)! (|V| - 1)!} \right. \\
&\quad \left. + \sum_{s \neq v} \left( \frac{k - |V|}{|V| - 1} + \sum_{\substack{p \in \partial_{sv} \\ |V| - \Psi(p) \geq k-1}} \frac{(|V| - \Psi(p))! (|V| - k)!}{\sigma_{sv} (|V| - \Psi(p) - k + 1)! (|V| - 1)!} \right) \right). \tag{5.10}
\end{aligned}$$

While the closed-form formulas (5.10) and (5.5) allow us to efficiently compute in polynomial time the Shapley value-based and Semivalue-based betweenness centralities, respectively, in the following section we introduce algorithms that perform this computation even more efficiently.

#### 5.4. Algorithms to Compute the Shapley Value and Semivalue based Betweenness Centralities

Generally speaking, given a graph  $G$ , computing the Shapley value takes  $O(2^{|V(G)|})$  time (to consider all subsets of  $V(G)$ ). To circumvent this major obstacle, we proposed formulas (5.5) and (5.10) to compute in polynomial time the Shapley Value and the Semivalue, respectively. To speed up the computation even further, we propose in this section four algorithms:

1. Algorithm **SVB** computes the *Shapley value*-based betweenness centrality given an *unweighted* graph;
2. Algorithm **SB** computes the *Semivalue*-based betweenness centrality given an *unweighted* graph;
3. Algorithm **WSVB** computes the *Shapley value*-based betweenness centrality given a *weighted* graph;
4. Algorithm **WSB** computes the *Semivalue*-based betweenness centrality given a *weighted* graph.

We also show in this section how the above algorithms can be easily adapted to work on directed graphs. Interestingly, we show that Algorithm **SVB** has the same complexity as the best known algorithm to compute the standard betweenness centrality (due to Brandes, 2001). Our algorithms are based on a general framework proposed in this chapter, which generalises Brandes' framework [Brandes, 2001].

This section is structured as follows. Section 5.4.1 introduces our generalisation of Brandes' framework for unweighted graphs. Building upon it, Sections 5.4.2 and 5.4.3 introduce Algorithm **SVB** and Algorithm **SB**, respectively. After that, in Section 5.4.4, we introduce our generalisation of Brandes' framework for weighted graphs. Finally, Sections 5.4.5 and 5.4.6 introduce Algorithm **WSVB** and Algorithm **WSVB**, respectively.

### 5.4.1. The Framework for Unweighed Graphs

In order to compute the standard betweenness centrality for all nodes, a naïve algorithm would first compute the number of shortest paths between all pairs of nodes, and then for each node  $v$  sum up all pair-dependencies, which are defined as  $\frac{\sigma_{st}(v)}{\sigma_{st}}$ . This process takes  $O(|V|^3)$  time. Brandes [2001] proposed an algorithm to improve this complexity by using some recursive relation. This algorithm runs in  $O(|V| \cdot |E|)$  time, and requires  $O(|V| + |E|)$  space.

Next, we propose a framework that generalises Brandes' algorithm. We start by defining pair-dependency as:

$$\delta_{s,t}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}} f_{\delta}(d(s, t)), \quad (5.11)$$

which is the positive contribution that all shortest paths between  $s$  and  $t$  make to the assessment of vertex  $v$ . The value of the function  $f_{\delta}$  depends solely on the distance between  $s$  and  $t$ . For instance, if we set  $f_{\delta} = \frac{1}{d(s,t)}$  we obtain the distance-scaled betweenness centrality introduced by Borgatti and Everett [2006].

Next, we define one-side dependency as:

$$\delta_{s,\cdot}(v) = \sum_{t \in V} \delta_{s,t}(v), \quad (5.12)$$

which is the positive contribution that all shortest paths starting in vertex  $s$  make to the assessment of vertex  $v$ . Now, building upon the recursive equation proposed by Brandes [2001], we obtain the following:

$$\delta_{s,\cdot}(v) = \sum_{\substack{w: (v,w) \in E \\ d(s,w)=d(s,v)+1}} \frac{\sigma_{sv}}{\sigma_{sw}} \left( f_{\delta}(d(s, w)) + \delta_{s,\cdot}(w) \right). \quad (5.13)$$

Using the above equation, we can compute what we call the *parameterised betweenness centrality*  $c_{fg}$  for a vertex  $v$  by iterating over all other vertices and summing their contributions as follows:

$$c_{fg}(v) = \sum_{s \neq v} \left( \delta_{s,\cdot}(v) + g_{\delta}(d(s, v)) \right), \quad (5.14)$$

where the value of the function  $g_{\delta}$  depends solely on the distance between  $s$  and  $v$ .

---

**Algorithm 11: PBC**—a general framework to compute the *parametrised betweenness centrality* in *unweighted* graphs

---

**Input:** Graph  $G = (V, E)$ , functions:  $f_\delta$  and  $g_\delta$   
**Data:** queue  $\mathcal{Q}$ , stack  $\mathcal{S}$  for each  $v \in V$  and some source  $s$   
 $d(s, v)$  : distance from  $v$  to the source  $s$   
 $Pred_s(v)$  : list of predecessors of  $v$  on the shortest paths from source  $s$   
 $\sigma_{sv}$  : the number of shortest paths from  $s$  to  $v$   
 $\delta_{s,\cdot}(v)$  : the one-side dependency of  $s$  on  $v$   
**Output:**  $c_{fg}(v)$  parametrised betweenness centrality for each vertex  $v \in V$

```

1 foreach  $v \in V$  do  $c_{fg}(v) \leftarrow 0$ ;
2 foreach  $s \in V$  do
3   foreach  $v \in V$  do
4      $Pred_s(v) \leftarrow \text{empty list}$ ;  $d(s, v) \leftarrow \infty$ ;  $\sigma_{sv} \leftarrow 0$ ;
5    $d(s, s) \leftarrow 1$ ;  $\sigma_{ss} \leftarrow 1$ ; enqueue  $s \rightarrow \mathcal{Q}$ ;
6   while  $\mathcal{Q}$  is not empty do
7     dequeue  $v \leftarrow \mathcal{Q}$ ; push  $v \rightarrow \mathcal{S}$ ;
8     foreach  $w$  such that  $(v, w) \in E$  do
9       if  $d(s, w) = \infty$  then
10         $d(s, w) \leftarrow d(s, v) + 1$ ; enqueue  $w \rightarrow \mathcal{Q}$ ;
11       if  $d(s, w) = d(s, v) + 1$  then
12         $\sigma_{sw} \leftarrow \sigma_{sw} + \sigma_{sv}$ ; append  $v \rightarrow Pred_s(w)$ ;
13   foreach  $v \in V$  do
14      $\delta_{s,\cdot}(v) \leftarrow 0$ ;
15   while  $\mathcal{S}$  is not empty do
16     pop  $w \leftarrow \mathcal{S}$ ;
17     foreach  $v \in Pred_s(w)$  do
18        $\delta_{s,\cdot}(v) \leftarrow \delta_{s,\cdot}(v) + \frac{\sigma_{sv}}{\sigma_{sw}} (f_\delta(d(s, w)) + \delta_{s,\cdot}(w))$ ;
19     if  $w \neq s$  then
20        $c_{fg}(w) \leftarrow c_{fg}(w) + \delta_{s,\cdot}(w) + 2g_\delta(d(s, w))$ ;
21 foreach  $v \in V$  do
22    $c_{fg}(v) = \frac{c_{fg}(v)}{2}$ ;

```

---

Now, we are ready to introduce our generalised framework, **PBC** (see Algorithm 11) for computing the *parametrised betweenness centrality* in unweighted graphs. More specifically, this framework modifies Brandes' approach and computes the betweenness centrality parametrised by two functions:  $f_\delta$  and  $g_\delta$ . All lines that differ from the original Brandes' algorithm were highlighted.

Firstly, in lines 6 - 12, the algorithm calculates both the distance and the number of shortest paths from a source  $s$  to each vertex. These lines are exactly the same as in the original

algorithm introduced by Brandes. While executing these lines, for each vertex  $v$ , all directly preceding vertices occurring on shortest paths from  $s$  to  $v$  are stored in memory. This process uses Breadth-First Search [Cormen, 2001] which takes  $O(|V|)$  time and  $O(|V| + |E|)$  space. In the second step (lines 18 and 20), the algorithm uses formula (5.14) to calculate the contribution of the source  $s$  to the value of our betweenness centrality for each vertex that is reachable from the source. This step also takes  $O(|V|)$  time and  $O(|V| + |E|)$  space. Only lines 18 and 20 and additionally line 21 differ from the original Brandes algorithm.

As visible in formula (5.14), in an undirected graph, each path is considered twice. Thus, at the end of the algorithm, in line 21, we halve the accumulated result. In line 20, we multiply the influence of the function  $g_\delta$  by two, because the influence of each source even in undirected graphs is considered only once. Finally, we note that it is very easy to adopt Algorithm 11 to directed graphs. To this end, we remove the loop from line 21 and halve the contribution of the vertex  $s$  from line 20, which now should look as follows:

$$\mathbf{20} : c_{fg}(w) \leftarrow c_{fg}(w) + \delta_{s,\cdot}(w) + g_\delta(d(s, w));$$

So, for directed and undirected graphs, the algorithm runs in  $O(|V| \cdot |E|)$  time, and requires  $O(|V| + |E|)$  space.

#### 5.4.2. The Algorithm for the Shapley Value-based Betweenness Centrality for Unweighted Graph

In this subsection we will construct an efficient algorithm for computing the Shapley value-based betweenness centrality for unweighted graphs. Specifically, in such graphs, the number of vertices on the shortest path between  $s$  and  $t$  is equal to the distance between  $s$  and  $t$ , denoted as  $d(s, t)$ .<sup>5</sup> In other words, we have:  $|\Psi(p)| = d(s, t)$ . Based on this, it is possible to simplify (5.5) as follows:

$$\begin{aligned} \phi_v^{SV}(V(G), \nu) &= \sum_{s \neq v \neq t} \sum_{p \in \partial_{st}(v)} \frac{1}{\sigma_{st} d(s, t)} + \sum_{s \neq v} \sum_{p \in \partial_{sv}} \frac{2 - d(s, v)}{2\sigma_{sv} d(s, v)} \\ &= \sum_{s \neq v} \left( \sum_{t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st} d(s, t)} + \frac{2 - d(s, v)}{2d(s, v)} \right). \end{aligned} \quad (5.15)$$

The above equation provides the same insight as equation (5.5) but for specific unweighted graphs. By transforming the second element of the inner sum  $\frac{2 - d(s, v)}{2d(s, v)} = \frac{1}{d(s, v)} - \frac{1}{2}$  we find that, in unweighted graphs, the Shapley value using group betweenness centrality as a characteristic function is in fact the sum of *the distanced scaled betweenness centrality* (introduced by

<sup>5</sup> For notational convenience, we assume the distance between two vertices to be the number of vertices on the shortest path between them (not the number of edges), e.g.,  $d(s, s) = 1$ .

---

**Algorithm 12: SVB SV-based betweenness centrality**

---

**Input:** Graph  $G = (V, E)$

**Output:**  $\phi_v^{SV}$  SV-based betweenness centrality for each vertex  $v \in V$

- 1  $f_\delta(d) \leftarrow \frac{1}{d}$ ;
  - 2  $g_\delta(d) \leftarrow \frac{2-d}{2d}$ ;
  - 3 **PBC**( $G, f_\delta, g_\delta$ )
- 

Borgatti and Everett, 2006) and the *closeness* centrality, shifted by half. Additionally, the above equation allows us to compute the Shapley value-based betweenness centrality in  $O(|V|^3)$ , but this result will be improved further.

Now, we adopt the framework presented in the previous subsection in order to accommodate equation (5.15). We simply need to set  $f_\delta = \frac{1}{d(s,t)}$  and set  $g_\delta = \frac{2-d(s,v)}{2d(s,v)}$ . This way, we can use our general framework—**PBC**, see Algorithm 11—to compute the Shapley value-based betweenness centrality in  $O(|V||E|)$ . This method is presented in Algorithm 12.

### 5.4.3. The Algorithm for the Semivalue-based Betweenness Centrality for Unweighted Graphs

We will follow the same reasoning as in the previous subsection. Specifically, since in unweighted graphs it holds that  $|\Psi(p)| = d(s, t)$ , we can transform equation (5.10) into:

$$\begin{aligned} \phi_v^{SEMI}(V(G), \nu) &= \sum_{1 \leq k \leq |V|} PD(V, k) \left( \sum_{s \neq v \neq t} \sum_{\substack{p \in \partial_{st}(v) \\ |V| - d(s,t) \geq k-1}} \frac{(|V| - d(s, t))! (|V| - k)!}{\sigma_{st} (|V| - d(s, t) - k + 1)! (|V| - 1)!} \right. \\ &\quad \left. + \sum_{s \neq v} \left( \frac{k - |V|}{|V| - 1} + \sum_{\substack{p \in \partial_{sv} \\ |V| - d(s,v) \geq k-1}} \frac{(|V| - d(s, v))! (|V| - k)!}{\sigma_{sv} (|V| - d(s, v) - k + 1)! (|V| - 1)!} \right) \right) \\ &= \sum_{1 \leq k \leq |V|} PD(V, k) \left( \sum_{\substack{s \neq v \neq t \\ |V| - d(s,t) \geq k-1}} \frac{\sigma_{st}(v) (|V| - d(s, t))! (|V| - k)!}{\sigma_{st} (|V| - d(s, t) - k + 1)! (|V| - 1)!} \right. \\ &\quad \left. + \sum_{\substack{s \neq v \\ |V| - d(s,v) \geq k-1}} \left( \frac{(|V| - d(s, v))! (|V| - k)!}{(|V| - d(s, v) - k + 1)! (|V| - 1)!} \right) + k - |V| \right). \end{aligned} \quad (5.16)$$

Our framework can be easily adopted to deal with the Semivalue; all we need to do is to set:

$$f_\delta = \begin{cases} \frac{(|V| - d)! (|V| - k)!}{(|V| - d - k + 1)! (|V| - 1)!} & \text{if } |V| - d(s, t) \geq k - 1 \\ 0 & \text{if } |V| - d(s, t) < k - 1 \end{cases}$$

---

**Algorithm 13: SB Semivalue-based betweenness centrality**

---

**Input:** Graph  $G = (V, E)$

**Output:**  $\phi_v^{SEMI}$  Semivalue-based betweenness centrality for each vertex  $v \in V$

```
1  $f_\delta(d) \leftarrow$  if  $|V| - d \geq k - 1$  then  $\frac{(|V|-d)! (|V|-k)!}{(|V|-d-k+1)! (|V|-1)!}$  else 0  $g_\delta(d) \leftarrow f_\delta(d) + \frac{k-|V|}{|V|-1};$ 
2 for  $k \leftarrow 1$  to  $|V|$  do
3   PBC( $G, f_\delta, g_\delta$ );
4   foreach  $v \in V$  do
5      $\phi_v^{SEMI} \leftarrow PD(V, k) c_{fg}(v);$ 
```

---

and set  $g_\delta = f_\delta + \frac{k-|V|}{|V|-1}$ . This way, we can use equation (5.16) along with our general framework—**PBC**, see Algorithm 11—to compute the Semivalue-based betweenness centrality for unweighted graphs in  $O(|V|^2|E|)$  time. The pseudo code is provided in Algorithm 13.

#### 5.4.4. The Framework for Weighted Graphs

While the focus of the previous subsections was on unweighted graphs, in this subsection we show how to generalise our framework to deal with weighted graphs. In particular, we consider one of the most popular semantics of weighted graphs, where the weight  $\lambda(v, u)$  of the edge between  $v$  and  $u$  is interpreted as the distance between  $v$  and  $u$ . Thus, unlike the case with unweighted graphs, where  $|\Psi(p)| = d(s, t)$ , in the case of weighted graphs this equality does not necessarily hold.

To introduce our generalised framework, we need additional notation. Let  $T_{st}[i] : i \in \{1, \dots, |V|\}$  be the number of shortest paths between  $s$  and  $t$  that contain exactly  $i$  vertices. The array  $T_{st}$  uniquely determines the polynomial  $W_{st}$  with terms  $T_{st}[i]x^i$ . We define the following operations on  $T_{st}$ :

- **Shifting:**  $T_{st}^{\rightarrow}$  and  $T_{st}^{\leftarrow}$  increase and decrease the indices of all values of the array by one, respectively. This takes  $O(|V|)$  time.
- **Adding:**  $T_{sv} \oplus T_{su}$  is the operation of adding two polynomials  $W_{sv}$  and  $W_{su}$ . It takes  $O(|V|)$  time. We will denote by  $\oplus$  the sum of a series of polynomials.
- **Multiplying:**  $T_{sv} \otimes T_{vt}$  is the operation of multiplying two polynomials  $W_{sv}$  and  $W_{vt}$ . This takes  $O(|V| \log |V|)$  time using the polynomial multiplying algorithm from [Cormen, 2001].
- **Resetting:**  $T_{sv} \leftarrow 0$  is an operation that assigns 0 to each cell in  $T_{sv}$ .

Next, we will show how the above operations allow us to tackle two algorithmic problems: (i) how to count all shortest paths and (ii) how to derive the recursive relation in order to compute one-side dependency. We start by considering the first of those problems, i.e., counting the

shortest paths. Here, we will use the following relation:

$$T_{sv} = \bigoplus_{u: d(s,u)+\lambda(u,v)=d(s,v)} T_{su}^{\rightarrow}. \quad (5.17)$$

Using Dijkstra's algorithm [Cormen, 2001], as well as equation (5.17), we can compute  $T_{st}$  for every pair of nodes  $s$  and  $t$ . If vertex  $u$  immediately precedes vertex  $v$  on some shortest path from source  $s$ , all shortest paths stored in  $T_{su}$  extended by vertex  $v$  are part of the set of shortest paths stored in  $T_{sv}$ . This procedure takes  $O(|V|^2|E| + |V|^2 \log |V|)$  time.

Secondly, in order to count all paths passing through a given node  $v$ , we will introduce the following relationship, where  $T_{st}(v)$  is an array defined just like  $T_{st}$  except that it only counts the shortest paths between vertices  $s$  and  $t$  that pass through  $v$ :

$$T_{st}(v) = (T_{sv} \otimes T_{vt})_{st}^{\leftarrow} = T_{sv} \otimes T_{vt}^{\leftarrow}. \quad (5.18)$$

The above equation can be interpreted as follows. Every path stored in the array  $T_{sv}$  can be extended by every path stored in the array  $T_{vt}$ . The outcome of this operation, which is in fact the multiplication of the two polynomials  $W_{sv}$  and  $W_{vt}$ , gives us information about all shortest paths from  $s$  to  $t$  passing through  $v$ . The vertex  $v$  is counted twice. To avoid duplicate counting, we shift the result of multiplication to the left, which effectively reduces the number of nodes in each path by one.

In weighted graphs the pair-dependency limited to the shortest paths consisting of  $i$  nodes is defined as follows:

$$\delta_{s,t}^*(v)[i] = \frac{T_{st}(v)[i]}{\sigma_{st}} f_{\delta^*}(i). \quad (5.19)$$

This limited pair-dependency measures the influence of all shortest paths between  $s$  and  $t$  consisting of  $i$  nodes on the evaluation of vertex  $v$ . The value of the function  $f_{\delta^*}$  depends solely on the number of nodes lying on the shortest paths between  $s$  and  $t$ .

Taking all possible values of  $i$  into consideration, we obtain the pair-dependency in weighted graphs:

$$\delta_{s,t}^*(v) = \sum_{i=1}^{|V|} \delta_{s,t}^*(v)[i], \quad (5.20)$$

which is the positive contribution that all shortest paths between  $s$  and  $t$  make to the assessment of vertex  $v$ ,

The definition of one-side dependency is similar to the one presented earlier in Subsection 5.4.1, except that we replace  $\delta$  with  $\delta^*$ :

$$\delta_{s,\cdot}^*(v) = \sum_{t \in V} \delta_{s,t}^*(v). \quad (5.21)$$

This is the positive contribution that all shortest paths starting with vertex  $s$  make to the evaluation of vertex  $v$ . The version of equation (5.21) that only considers the shortest paths containing exactly  $i$  nodes each is:

$$\delta_{s,\cdot}^*(v)[i] = \sum_{t \in V} \delta_{s,t}^*(v)[i], \quad (5.22)$$

Now, we are able to infer the recursive relation for weighted graphs:

$$\delta_{s,\cdot}^*(v)[i] = \sum_{\substack{w: (v,w) \in E \\ d(s,w)=d(s,v)+1}} \frac{T_{sv}[i]}{\sigma_{sw}} \left( f_{\delta^*}(i+1) + \delta_{s,\cdot}^*(w)[i+1] \right). \quad (5.23)$$

Using the above equation, we are able to compute a parametrised betweenness centrality for weighted graphs  $c_{fg}^*$ . More specifically, for a vertex  $v$ , we compute  $c_{fg}^*(v)$  by iterating over all other vertices and summing their contributions. More formally:

$$c_{fg}^*(v) = \sum_{s \neq v} \sum_{i=1}^{|V|} \left( \delta_{s,\cdot}^*(v)[i] + \frac{T_{sv}[i]}{\sigma_{sv}} g_{\delta^*}(i) \right), \quad (5.24)$$

where the value of the function  $g_{\delta^*}$  depends solely on the number of nodes lying on the shortest paths between  $s$  and  $v$ .

Now, we are ready to introduce our general framework, namely **WPBC** (see Algorithm 14) to compute the *parametrised betweenness centrality* in weighted graphs. All lines that differ from the original Brandes' algorithm were highlighted. In lines 3 - 12 the algorithm uses Dijkstra's algorithm [Cormen, 2001] to traverse the graph and count all shortest paths from the source  $s$ . In these lines it essentially differs from Brandes algorithm, because it uses polynomial arithmetic to count paths and number of nodes lying on these paths. Next, in lines 18 and 20, which also differ from Brandes' framework, we implemented the recursive formula (5.23).

Algorithm 14 runs in  $O(|V|^2|E| + |V|^2 \log |V|)$  time and requires  $O(|V|^2)$  space. Furthermore, this algorithm (just like Algorithm 11) can be easily adapted to directed graphs.

#### 5.4.5. The Algorithm for Shapley Value-based Betweenness Centrality in Weighted Graphs

In this subsection we will construct an efficient algorithm for computing the Shapley Value-based betweenness centrality for weighted graphs. We will use the notation introduced in the previous subsection, and transform equation (5.5) into:



---

**Algorithm 14: WPBC**—a general framework to compute the *parametrised betweenness centrality* in *weighted graphs*

---

**Input:** weighted graph  $G = (V, E)$ , with weight function  $\lambda : E \rightarrow \mathbb{R}^+$

**Data:** priority queue  $\mathcal{Q}$  with key  $d()$ , stack  $\mathcal{S}$

$d(s, v)$  : the distance from  $s$  to  $v$

$Pred_s(v)$  : the list of predecessors of  $v$  on the shortest paths from source  $s$

$\sigma_{sv}$  : the number of shortest paths from  $s$  to  $v$

$\delta_{s,\cdot}^*(v)[i]$  : one-side dependency of  $s$  on  $v$

$T[i]_{sv}$  : the number of shortest paths from  $s$  to  $v$  containing  $i$  vertices

**Output:**  $c_{fg}^*(v)$  parametrised betweenness centrality

```

1 foreach  $s \in V$  do
2   foreach  $v \in V$  do  $Pred_s(v) \leftarrow \text{empty list}; d(s, v) \leftarrow \infty; \sigma_{sv} \leftarrow 0;$ 
    $d(s, s) \leftarrow 1; \sigma_{ss} \leftarrow 1; T_{ss}[1] \leftarrow 1; \text{enqueue } s \rightarrow \mathcal{Q};$ 
3   while  $\mathcal{Q}$  is not empty do
4      $\text{extract } v \leftarrow \mathcal{Q}$  with minimal  $d(s, v); \text{push } v \rightarrow \mathcal{S};$ 
5     foreach  $w$  such that  $(v, w) \in E$  do
6       if  $d(s, w) > d(s, v) + \lambda(v, w)$  then
7          $d(s, w) \leftarrow d(s, v) + \lambda(v, w);$ 
8          $\text{insert/update } w \rightarrow \mathcal{Q}$  with  $d(s, w);$ 
9          $\sigma_{sw} \leftarrow 0; T_{sw} \leftarrow 0; Pred_s(w) \leftarrow \text{empty list};$ 
10      if  $d(s, w) = d(s, v) + \lambda(v, w)$  then
11         $\sigma_{sw} \leftarrow \sigma_{sw} + \sigma_{sv}; T_{sw} = T_{sw} \oplus T_{sv}^{\rightarrow};$ 
12       $\text{append } v \rightarrow Pred_s(w);$ 
13   foreach  $v \in V$  do
14      $\delta_{s,\cdot}^*(v) \leftarrow 0;$ 
15   while  $\mathcal{S}$  is not empty do
16      $\text{pop } w \leftarrow \mathcal{S};$ 
17     foreach  $v \in Pred_s(w)$  do for  $i \leftarrow 1$  to  $|V| - 1$  do
18        $\delta_{s,\cdot}^*(v)[i] \leftarrow \delta_{s,\cdot}^*(v)[i] + \frac{T_{sv}[i]}{\sigma_{sw}} \left( f_{\delta^*}(i+1) + \delta_{s,\cdot}^*(w)[i+1] \right);$ 
19     if  $w \neq s$  then for  $i \leftarrow 1$  to  $|V|$  do
20        $c_{fg}^*(w) \leftarrow c_{fg}^*(w) + \delta_{s,\cdot}^*(w)[i] + 2 \frac{T_{sw}[i]}{\sigma_{sw}} g_{\delta^*}^*(i);$ 
21 foreach  $v \in V$  do  $c_{fg}^*(v) = \frac{c_{fg}^*(v)}{2};$ 

```

---

---

**Algorithm 15: WSVB Weighted SV-based betweenness centrality**


---

**Input:** weighted graph  $G = (V, E)$ , with weight function  $\lambda : E \rightarrow \mathbb{R}^+$

**Output:**  $\phi_v^{SV}$  weighted SV-based betweenness centrality for each  $v \in V$

- 1  $f_\delta^*(i) \leftarrow \frac{1}{i}$ ;
  - 2  $g_\delta^*(i) \leftarrow \frac{2-i}{2i}$ ;
  - 3 **WPBC**( $G, \lambda, f_\delta^*, g_\delta^*$ )
- 

$$\phi_v^{SV}(V(G), \nu) = \sum_{s \neq v} \left( \sum_{t \neq v} \sum_{i=1}^{|V|} \frac{T_{st}(v)[i]}{\sigma_{st} i} + \sum_{i=1}^{|V|} \frac{T_{sv}[i](2-i)}{2\sigma_{sv} i} \right). \quad (5.25)$$

The framework for weighted graphs (introduced in Section 5.4.4) can be used to compute (5.25). All we need is to define  $f_\delta^*(i) = \frac{1}{i}$  and  $g_\delta^*(i) = \frac{2-i}{2i}$ . By doing so, we can use our general framework—**WPBC**, see Algorithm 14—to compute the weighted Shapley value-based betweenness centrality for all nodes in  $O(|V|^2|E| + |V|^2 \log |V|)$ . This method is presented in Algorithm 15.

#### 5.4.6. The Algorithm for Semivalue-based Betweenness Centrality in Weighted Graphs

In this subsection we use our framework for weighted graphs, namely **WPBC** (see Algorithm 14), to compute the Semivalue-based betweenness centrality for weighted graphs. To this end, we can transform equation (5.10) into:

$$\begin{aligned} \phi_v^{SEMI}(V(G), \nu) = \sum_{1 \leq k \leq |V|} PD(V, k) & \left( \sum_{s \neq v} \sum_{t \neq v} \sum_{i=1}^{|V|-i \geq k-1} \frac{T_{st}(v)[i](|V|-i)!(|V|-k)!}{\sigma_{st}(|V|-i-k+1)!(|V|-1)!} \right. \\ & \left. + \sum_{s \neq v} \left( \frac{k-|V|}{|V|-1} + \sum_{i=1}^{|V|-i \geq k-1} \frac{T_{sv}[i](|V|-i)!(|V|-k)!}{\sigma_{sv}(|V|-i-k+1)!(|V|-1)!} \right) \right). \end{aligned} \quad (5.26)$$

Our framework can be easily adopted to deal with Semivalues; all we need is to set:

$$f_\delta^*(i) = \begin{cases} \frac{(|V|-i)!(|V|-k)!}{(|V|-i-k+1)!(|V|-1)!} & \text{if } |V| - i \geq k - 1 \\ 0 & \text{if } |V| - i < k - 1 \end{cases}$$

and set  $g_\delta^* = f_\delta^* + \frac{k-|V|}{|V|-1}$ . This way, we can use our general framework for weighted graphs, namely **WPBC** (see Algorithm 14) to compute the Semivalue-based betweenness centrality in  $O(|V|^3|E| + |V|^3 \log |V|)$  time. The pseudo code is presented in Algorithm 16.

---

**Algorithm 16: WSB Weighted Semivalue-based betweenness centrality**

---

**Input:** weighted graph  $G = (V, E)$ , with weight function  $\lambda : E \rightarrow \mathbb{R}^+$

**Output:**  $\phi_v^{SEMI}$  weighted Semivalue-based betweenness centrality for  $v \in V$

```
1  $f_\delta^*(i) \leftarrow$  if  $|V| - i \geq k - 1$  then  $\frac{(|V|-i)! (|V|-k)!}{(|V|-i-k+1)! (|V|-1)!}$  else 0  $g_\delta^*(i) \leftarrow f_\delta^*(i) + \frac{k-|V|}{|V|-1};$   
2 for  $k \leftarrow 1$  to  $|V|$  do  
3   WPBC( $G, \lambda, f_\delta^*, g_\delta^*$ );  
4   foreach  $v \in V$  do  
5      $\phi_v^{SEMI} \leftarrow PD(V, k) c_{fg}^*(v);$ 
```

---

## 5.5. Simulations

So far, we argued that centrality measures based on Semivalues in general, and on the Shapley value in particular, can help analyse networks in situations where simultaneous incidents may occur. Furthermore, we presented polynomial time algorithms to compute both the Shapley value-based and Semivalue-based betweenness centralities, for weighted graphs as well as unweighted graphs. This section provides an empirical evaluation of the above contributions. In particular, Section 5.5.1 compares our centrality measures against standard betweenness centrality, in a scenario where simultaneous node failures are simulated. In particular, we compare the effectiveness of both measures when tackling Problem 3. Section 5.5.2 evaluates the running time of our algorithms, and compares them against the Monte Carlo method—the only available alternative in the literature.

We carry out our experiments on weighted and unweighted random scale-free graphs, which are created using the *preferential attachment* mechanism introduced by Barabási and Albert [1999]. We also carry out experiments on two real-life network: (1) the email communication networks introduced by Guimerà et al. [2003], which contains 1133 nodes and 5451 edges, and (2) the neural system of *C. elegans* studied by Watts and Strogatz [1998], which contains 297 nodes and 2359 edges.

### 5.5.1. Evaluating our Centrality Measure

In this subsection, we compare the node ranking obtained by the Semivalue-based betweenness centrality against the node ranking obtained by standard betweenness centrality. Note that, by this evaluation, we also implicitly evaluate our Shapley value-based betweenness centrality, as it is a special case of the Semivalue-based one.

To this end, let us consider the scenario of network vulnerability (see Section 5.3.1 for more details). The comprehensive work by Holme et al. [2002] examines how different strategies of attack on the network affect its performance. In more detail, the authors studied how removing the nodes with the highest betweenness centrality affects communication throughout

the network. They found that, in random scale-free networks, removing the nodes with the highest betweenness centrality causes an exponential decay in network performance. Here, performance is measured as the average distance between nodes—an important measure used, e.g., in the seminal paper of Albert et al. [2000] on network vulnerability to attacks and random node failures.

The literature closely related to the work done by Holme et al. [2002] is about network vulnerability analysis based on three-stage Stackelberg games [Stackelberg, 1952]. In a nutshell, two opponents, an *interdictor* and an *operator*, try to maximize their gains. In the first stage, the operator fortifies the network by choosing the nodes that should be fully protect. Next, the interdictor performs an attack, which involves removing a limited number of unfortified nodes. Finally, the operator acts on the network in order to fulfill its objectives, e.g., he can travel between two nodes, or send some load between nodes. The interdictor can be human or can be of a different nature such as a natural disaster or a random failure of nodes. In such cases one can assume worst-case scenarios (Murphy’s Law) [Smith and Lim, 2008], or some stochastic model of interdictor attacks [Cormican et al., 1998]. The instance of Stackelberg games closely related to this problem is called *Shortest Path Interdiction*. In such games the *interdictor* attempts to remove  $k$  edges in order to maximize the distance between two nodes  $s$  and  $t$ . This problem is proven to be NP-hard [Ball et al., 1989]. However, Malik et al. [1989] proposed an effective approximation algorithm and Corley and Sha [1982] solved this problem in polynomial time for the special case where  $k = 1$ .

In this subsection we follow the Holme’s line of research. We study the resilience of a network against random failures, where we assume that the level of protection of each node is determined based on its ranking. In order to carry out our experiment, we first need to agree on the way in which the functionality of the network will be measured. To this end, we use the average inverse geodesic measure (IGM), which was used by Albert et al. [2000] and Holme et al. [2002]. This measure captures the notion of average distance between nodes in a network, and is given by the following formula:

$$\text{IGM}(V, E) = \sum_{v \in V} \sum_{v \neq u \in V} \frac{1}{d(v, u)}.$$

Our experiment consists of two phases. The first involves generating random graphs and computing node cardinal rankings. This phase involves the following steps:

1. Generate an unweighted random scale-free graph containing 160 nodes, in which the average node degree is  $\sqrt{160}$ .<sup>6</sup>
2. Set the interval  $[a, b)$ , which means that the number of nodes that will be simultaneously exposed to failure is between  $a$  (inclusive) and  $b$  (exclusive).

---

<sup>6</sup> We also experimented with alternative values; the results exhibited similar trends.

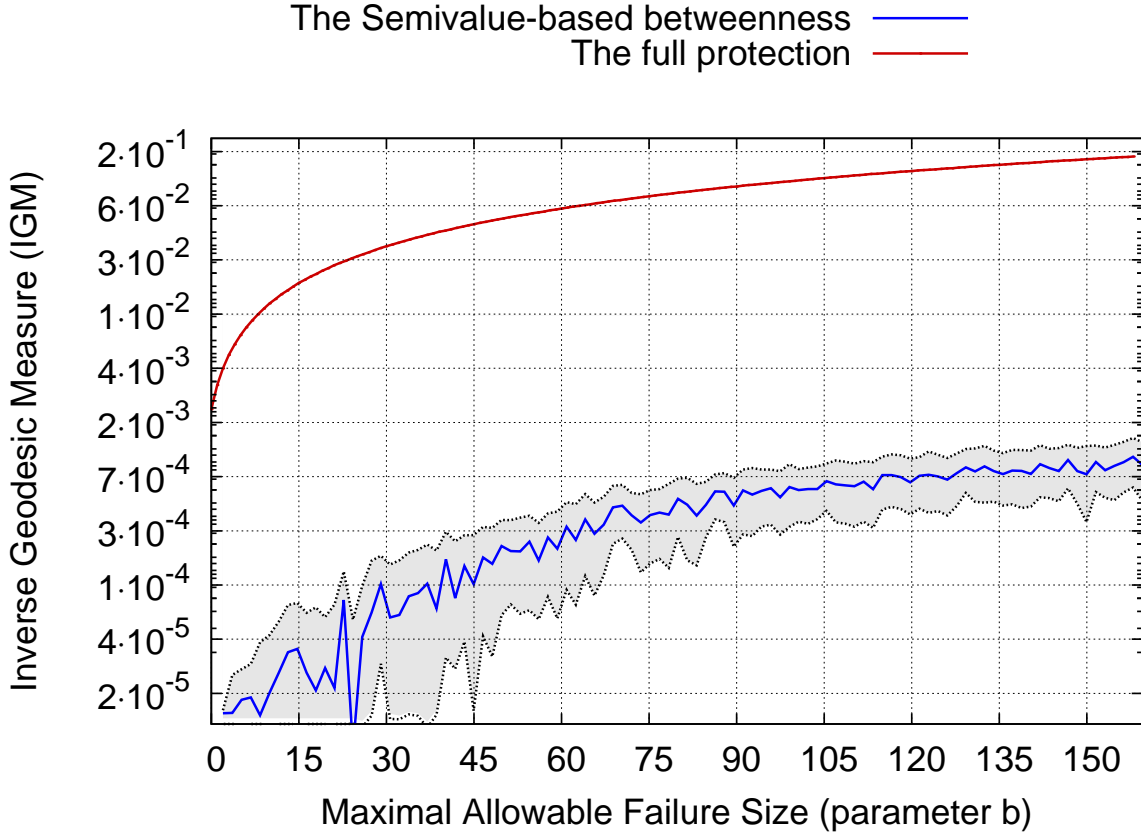


Figure 5.3: The blue line plots the difference in performance between the Semivalue-based measure and the standard betweenness centrality on random graphs, where  $a = 1$  and  $b = 2, \dots, 160$ , and where the protection strategy is set to protect nodes with probability  $\frac{1}{r(v)^2}$ . The confidence interval is set to 75%. The rightmost point of the plot (where the interval is  $[1, 160)$ ) represents the special case where the Semivalue-based measure happens to be identical to the Shapely value-based measure. The red line depicts the performance of the full protection strategy. The y-axis is in logscale.

3. Compute two rankings of nodes. The first is based on standard betweenness centrality. The second is based on Semivalue-based betweenness centrality with probability distribution:

$$PD(V, k) = \begin{cases} \frac{1}{b-a} & \text{if } k \in [a, b) \\ 0 & \text{otherwise.} \end{cases}$$

The second phase of the experiment involves the following steps:

1. Protect each node  $v \in V$ . The level of protection of each node is determined based on its position in the ranking, e.g., if a node is ranked first, then it is the most important node and so will have the highest level of protection. We study two approaches to nodes protection. The first one assigns to each node in the ranking the specific power of protection. More formally, if we denote by  $r(v)$  the position of node  $v$  in ranking  $r$ , then node  $v$  is protected with probability  $\frac{1}{r(v)^2}$ . The second one is to fully protect the top 10% of nodes in the ranking.

2. Choose randomly a set of nodes  $S \subseteq V$ , which will be exposed to failure.<sup>7</sup> Here, the number of nodes that are exposed to failure is chosen uniformly at random from the range  $[a, b)$ . Likewise, out of all subsets that have exactly this number of nodes, the subset  $S$  is chosen uniformly at random.
3. Using the first approach implies that  $v$  will fail with probability  $1 - \frac{1}{r(v)^2}$ . Using the second one indicates that each node  $v \in S$  will fail if  $r(v) < 0.1 * |V|$ .

Having explained each phase of the experiment, we now explain how the entire experiment works:

1. Generate 30 random graphs.
2. Set the size of the failure to  $[a, b)$ , where  $a = 1$  and  $b$  is running from 2 to 160.
3. For each  $[a, b)$ , simulate 5000 simultaneous node failures.
4. Protect nodes depending on the two rankings: standard betweenness centrality and Semivalue-based betweenness centrality.
5. Compare the average IGM that results from each measure together with confidence interval of 75%.
6. Compare results to the full protection strategy that will protect all nodes from being removed.

Figure 5.3 depicts the average difference in IGM between the two centrality measures. On the plot the blue line is  $IGM_S - IGM_B$ , where  $IGM_S$  is the functionality of the network when protected according to our centrality measure, while  $IGM_B$  is the functionality of the network when protected according to the standard betweenness centrality. The protection strategy is set to protect each node with probability  $\frac{1}{r(v)^2}$ . The values below zero indicate that standard betweenness centrality gives better protection to the network, whereas those above zero indicate that our Semivalue-based centrality gives better protection. As can be seen, when  $b$  is small, the performance of the Semivalue-based betweenness centrality is almost identical to the standard betweenness centrality. This is expected because our centrality measure, given the failure interval  $[1, 2)$ , is exactly the same as the standard measure. However, as we increase  $b$ , more nodes can fail simultaneously, and so the difference between our measure and the standard one become more evident. As can be seen, in such cases, our measure outperforms the standard one in terms of network protection, as more influential nodes (in terms of maintaining a low average distance throughout the network) are preserved. Note that for the interval  $[1, 160)$  the Semivalue-based centrality is exactly the same as the Shapley value-based

---

<sup>7</sup> In our setting, being *exposed* to failure does not necessarily imply that the node will fail; this depends on the level of protection that this node has against failure.

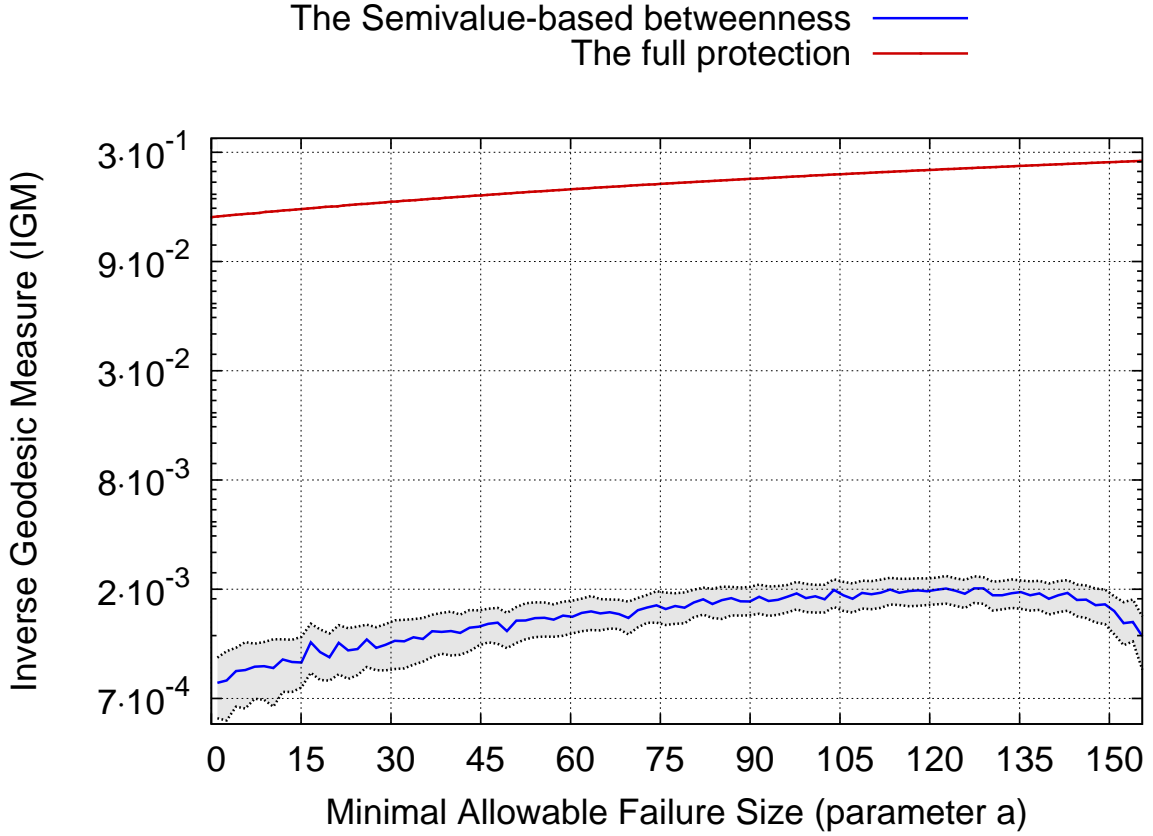


Figure 5.4: The blue line is the difference in performance between the Semivalue-based measure and the standard betweenness centrality on random graphs, where  $a = 1, \dots, 159$  and  $b = 160$ , and where the protection strategy is set to protect nodes with probability  $\frac{1}{r(v)^2}$ . The confidence interval is set to 75%. The leftmost point of the plot (where the interval is  $[1, 160)$ ) represents the special case where the Semivalue-based measure happens to be identical to the Shapely value-based measure. The red line depicts the performance of the full protection strategy. The y-axis is in logscale.

centrality. Additionally, the red line in Figure 5.3 depicts the performance of the full protection strategy that protects each node from being removed. It is  $IGM_O - IGM_B$ , where  $IGM_O$  is the functionality of the network when protected according to the full protection strategy, and  $IGM_B$  is the functionality of the network when protected according to the standard betweenness centrality.

Figure 5.4 depicts the same experiment, except that  $b = 160$  is now constant and  $a = 1, \dots, 159$ . As can be seen, for all values of  $a$  our measure gives better protection. Again note that for the interval  $[1, 160)$  the Semivalue-based centrality is exactly the same as the Shapely value-based centrality.

Figures 5.5 and 5.6 present the repetition of the above two experiments but for the other protection strategy, where we fully protect top 10% of nodes. The results are somewhat similar, but the advantage of our measure is less evident. For instance, given  $a = 30$  in Figure 5.6, the advantage of our measure is below  $7 \cdot 10^{-5}$ , whereas for the first protection strategy it

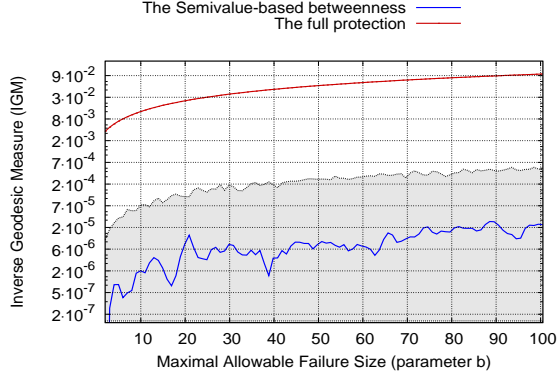


Figure 5.5: The difference in performance between the Semivalue-based measure and the standard betweenness centrality, where  $a = 1$  and  $b = 2, \dots, 100$ , and the strategy is to fully protect the top 10% of nodes.

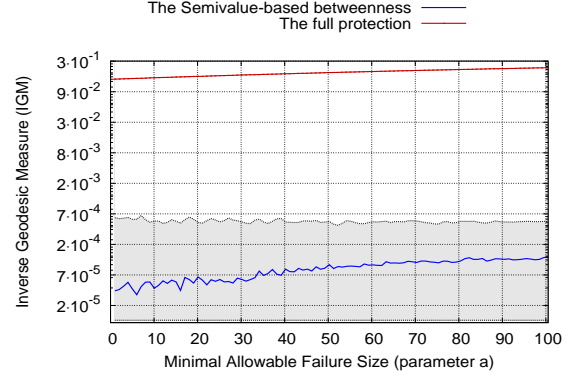


Figure 5.6: The difference in performance between the Semivalue-based measure and the standard betweenness centrality, where  $a = 1, \dots, 99$  and  $b = 100$ , and the strategy is to fully protect the top 10% of nodes.

is above  $7 \cdot 10^{-4}$ . This stems from the fact that the top nodes of the two rankings (i.e., the Semivalue-based and standard betweenness) are less different than the middle nodes or the bottom nodes. This is the consequence of the scale-free distribution and existence of powerful *hub* nodes that dominate two rankings. Thus, if we protect only the top 10% of nodes, the difference in performance will be less visible, but our new measure still outperforms standard betweenness centrality.

Finally, Figure 5.7 demonstrates that the performance between two specific Semivalue-based centralities: (1) the Shapley value-based and (2) the Banzhaf-based does not differ much. That is to say, the two measures can be used interchangeably in this application.

### 5.5.2. Evaluating our Algorithms

In this section, the running time of our exact algorithms is compared against that of the Monte Carlo (MC) sampling algorithm for computing the Shapley value.<sup>8</sup> Note that, due to the exponential number of coalitions in a coalitional game, sampling (as in the MC algorithm) is the only general-purpose method available in the literature for computing the Shapley value in reasonable time. Generally speaking, the MC algorithm works by iteratively sampling permutations of players (or nodes in our case). For each such permutation,  $\pi$ , and each node  $v$ , the algorithm computes the marginal contribution of  $v$  to the coalition consisting of every node that precedes  $v$  in  $\pi$ . Finally, after several iterations, the algorithm divides the aggregate sum of

<sup>8</sup> For a formalisation and a discussion of this algorithm, see [Castro et al., 2009] and [Michalak et al., 2013a], respectively.



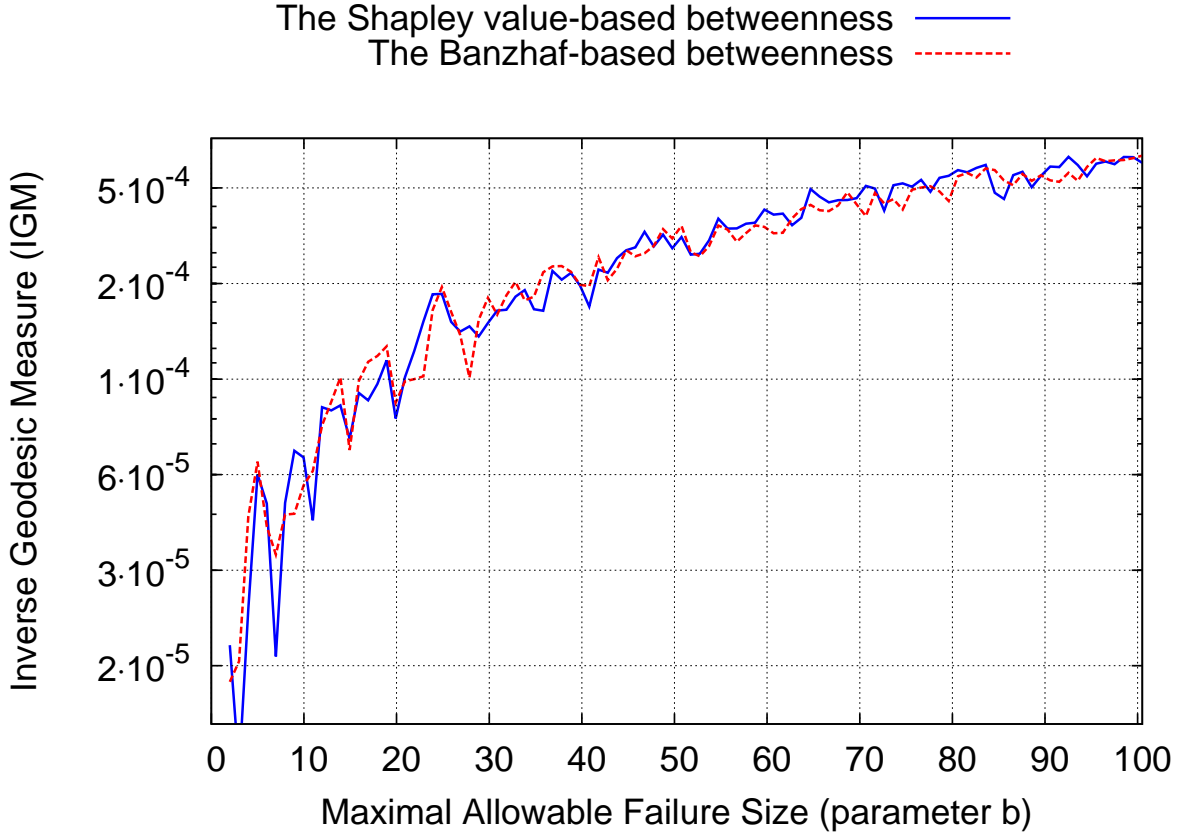


Figure 5.7: The blue line plots the difference in performance between the Shapley value-based measure and the standard betweenness centrality, and the red one the difference in performance between the Banzhaf-based measure and the standard betweenness centrality on random graphs, where  $a = 1$  and  $b = 2, \dots, 100$ , and where the protection strategy is set to protect nodes with probability  $\frac{1}{r(v)^2}$ . The y-axis is in logscale.

all marginal contributions for each node by the the number of iterations performed. The time complexity of this algorithm is  $O(Iter \times con)$ , where  $Iter$  denotes the number of iterations and  $con$  denotes the number of operations necessary for computing each marginal contribution. To optimise  $con$ , we implement MC sampling using the fast algorithm for computing group betweenness centrality presented by Puzis et al. [2007b]. This method guarantees that each MC iteration takes  $O(|V|^3)$  time.<sup>9</sup> Clearly, the MC sampling is never guaranteed to yield the exact Shapley values and its performance should be measured by an approximation error. In our simulations, we compute the average relative error ( $|1 - \frac{\text{approx. value}}{\text{exact value}}|$ ) for the value of each node in the network.

Starting with unweighted graphs, Figure 5.8 and Figure 5.9 compare the running time of our exact algorithms against the MC sampling given random networks, and given a real communication network, respectively. In more detail, Figures 5.8 depicts the average result

<sup>9</sup> Note that if we have oracle access to the coalition values (which is a common assumption in the literature), then  $con = 1$ . However, in our setting we do not make such an assumption. Instead, to obtain the value of a coalition, we need to compute its group betweenness centrality.

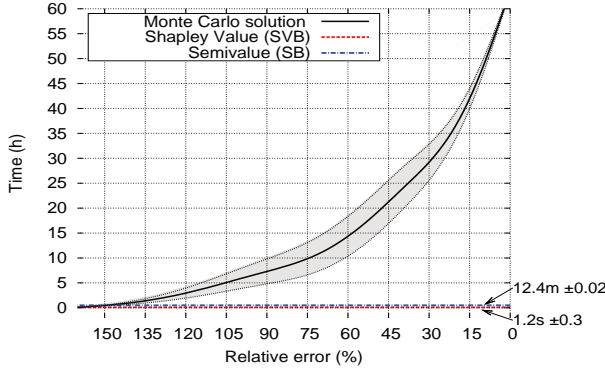


Figure 5.8: Random graphs with 500 nodes.

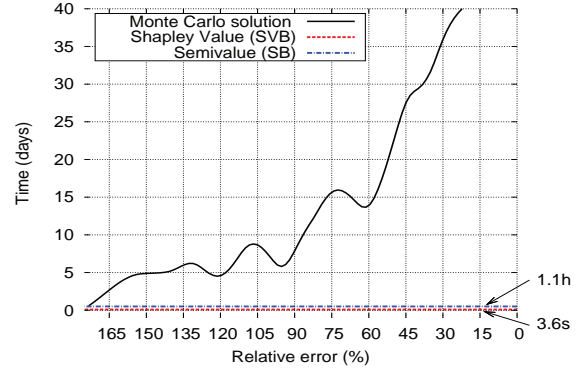


Figure 5.9: The communication network.

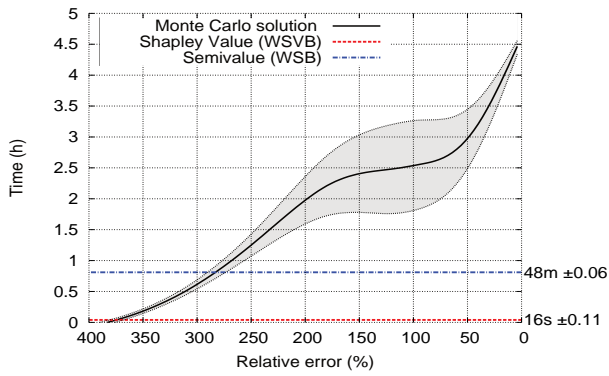


Figure 5.10: Random graphs with 200 nodes.

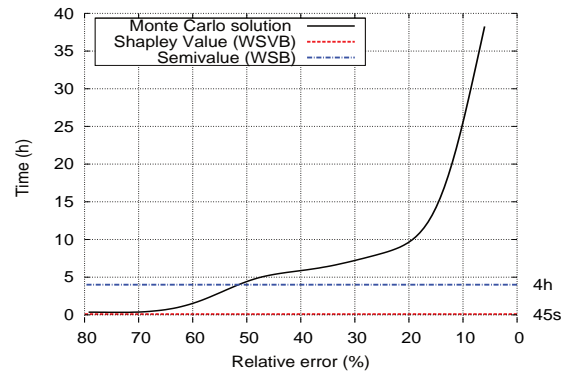


Figure 5.11: The neural system network.

taken over 30 randomly-generated graphs consisting of 500 nodes each, with an average degree of 25. The shaded area depicts the confidence interval of 75%. As can be seen, the Shapley value-based betweenness centrality is computed (using Algorithm 12) in 1.2 seconds, while and Semivalue-based betweenness centrality is computed (using Algorithm 13) in 12.4 minutes. On the other hand, in order to obtain an admissible average error of 10%, the MC sampling method takes around 45 hours. Furthermore, obtaining smaller average errors requires an exponential number of additional operations. Moving on to Figure 5.9, it depicts the results of the experiment performed on a real email communication network [Guimerà et al., 2003]. As can be seen, Algorithms 12 and 13 take 3.6 seconds and 1.1 hour to compute Shapley value-based and Semivalue-based betweenness centrality, respectively. In contrast, the MC sampling is unable to guarantee a better approximation than 15% even after 40 hours.

Having presented our results for unweighted graphs, we now move on to weighted graphs. Figures 5.10 and 5.11 compare the performance of our algorithms against the MC sampling, given random networks, and a real neural network, respectively. In more detail, Figure 5.10 depicts the average result taken over 30 randomly-generated graphs consisting of 200 nodes each, with an average degree of 15. Weights of edges were chosen uniformly at random from

Standard centrality	Group centrality	SV-based centrality	Efficient computation	Semivalue-based centrality	Efficient computation
<i>node</i>	Everett and Borgatti [1999]	Narayanam and Narahari (2010)	<b>Chapter 4</b>	[Szczepański et al., 2015b]	
<i>closeness</i>	Everett and Borgatti [1999]	<b>Chapter 4</b>	<b>Chapter 4</b>	[Szczepański et al., 2015b]	
<i>betweenness</i>	Everett and Borgatti [1999]	<b>this chapter</b>	<b>this chapter</b>	<b>this chapter</b>	<b>this chapter</b>

Table 5.2: Summary of the results obtained in this chapter vs. other contributions in the literature.

the range  $[1, 100]$ . The shaded area depicts the confidence interval of 75%. As can be seen, the Shapley value-based betweenness centrality is computed (using Algorithm 15) in 16 seconds, while the Semivalue-based betweenness centrality is computed (using Algorithm 16) in 48 minutes. On the other hand, in order to obtain an admissible average error of 10%, the MC sampling takes around 4 hours. Moving on to Figure 5.11, it depicts the experiment performed on a real neural network [Watts and Strogatz, 1998]. As can be seen, Algorithms 15 and 16 take 45 seconds and 4 hours to compute Shaple-value-based and Semivalue-based betweenness centralities, respectively.

## 5.6. Conclusions

In this chapter, was proposed an extension of betweenness centrality, based on Semivalues—a wide and flexible family of solution concepts from coalitional game theory. Our new measure provides a ranking of individual nodes while taking into consideration the synergies that may exist in different groups of nodes. The main contributions of this chapter are summarised in contrast to the literature in Table 5.2.

Despite the fact that Semivalues are, in general, computationally challenging, we showed that the new measure can be computed efficiently. More specifically, we proposed polynomial-time algorithms to compute all Semivalue-based betweenness centralities for weighted graphs as well as unweighted graphs. These include both the Shapley value-based and Banzhaf power index-based betweenness centralities. Interestingly, our algorithm for computing the Shapley value-based centrality for the unweighted networks has the same time complexity as the best known algorithm due to [Brandes, 2001] for computing the standard betweenness centrality.

We evaluated our algorithms empirically on weighted and unweighted random scale-free graphs Barabási and Albert [1999] as well as on two real networks. Our algorithms turned

out to be significantly faster than the Monte Carlo sampling [Castro et al., 2009]—the fastest alternative from the literature.

We also evaluated our centrality measures in a simulated scenario in which simultaneous node failures occur. Here, following Albert et al., 2000, and Holme et al., 2002, we quantified the functionality of the network based on the average inverse geodesic measure. The results showed that, compared to the standard measure, the ranking obtained by our game-theoretic network centrality reflects more accurately the influence that different nodes have on the network functionality.

## 6. A Centrality Measure for Networks With Community Structure Based on a Generalization of the Owen Value

In the previous chapters we presented the game-theoretic centrality measures based on Shaply value and Semivalues. In this chapter we propose the first measure of node centrality that takes into account the *community structure* of the underlying network. To allow for flexible modeling of community structures, we propose a generalization of the Owen value—a well-known solution concept from cooperative game theory to study games with *a priori*-given unions of players. As a result we obtain the first measure of centrality that accounts for both the value of an individual node’s relationships within the network and the quality of the community this node belongs to. We use our measure to evaluate citation networks, where all publications belong to naturally defined science communities united under journal titles and conference venues.

All technical contributions presented in this chapter were made exclusively by the author of this dissertation.

### 6.1. Introduction

Real-world networks frequently have highly complex structures. They can often be characterized by properties such as heavy-tailed degree distributions, clustering, the small-world property, etc. Another important characteristics that many real-life networks have in common is their *community structure* [Girvan and Newman, 2002, Newman, 2006] (see Section 2.8). Communities are usually composed of nodes that are more densely connected internally than with other nodes in the network. For instance, the teachers from a particular secondary school may form a community within the social network of all teachers in the city. Similarly, trade links among the European Union countries are usually more intense than their links with the rest of the world. In addition, certain communities may be considered to be stronger than others. Secondary schools may vary in reputation, and some trade blocks may be more important to the global economy than others.

The importance of a community is usually increased when a new, powerful individual joins it. Conversely, membership in a strong community may boost the importance of an otherwise weak individual. Quantifying this latter effect is the primary goal of this paper. In other words,

we are concerned with the problem of analysing the importance (the *centrality*) of individual nodes given the underlying community structure of the network.

In this chapter, we use the flexibility offered by the game-theoretic approach to centrality measures and constructed first centrality measure in the literature that is able to account for complex community structures in networks. To this end, we model the community structure as the *a priori* given coalition structure of a cooperative game. By doing so, we are able to build a centrality metric by generalizing the Owen value [Owen, 1977]—a well-known solution concept for cooperative games in which players are partitioned into pre-defined groups.

In our approach, the computation of a node’s power is a two-step process. First, we compute the importance of the community (if any) that this node belongs to. Next, we compute the power of the given node within this community. Our generalization of the Owen value, which we call *coalitional semivalues*, is a much broader solution concept. In fact, coalitional semivalues encompass the Owen value as well as all other solution concepts in the literature that were developed for games with an *a priori* defined coalition structure of players: the Owen-Banzhaf value [Owen, 1982], the symmetric coalitional Banzhaf value, and p-binomial semivalues [Carreras and Puente, 2012].

Although, in general, the new centrality introduced in this chapter is #P-complete (and hence NP-hard), we are able to give a polynomial algorithm to compute it for problem instances where the value of any group of nodes is determined by their degree centrality [Freeman, 1979] (see Definition 15). We verify the practical aspects of our algorithm on a large citation network that contains more than 2 million nodes and links. Our experiments compare three different degree centralities: group degree centrality [Everett and Borgatti, 1999], the Shapley value-based degree centrality (see Definition 26), and our new centrality. We show that, unlike others, our new centrality produces a ranking in which the power of the top nodes significantly differs, depending on the power of the communities that these nodes belong to.

## 6.2. Preliminaries

In this section we introduce the basic notation for coalitional games with coalitional structure and also we present the Owen value [Owen, 1977]—the most popular extension of the Shapley value to games with coalitional structure.

### 6.2.1. The Coalitional Structure

We already know that the *coalition* is a group of nodes in coalitional game. The coalitions can be non-overlapping, which mean that player can belong to only one coalition, and overlapping. In this dissertation we focus on non-overlapping coalitions. The game divided into fixed coalitions forms a *coalitional structure*.

**Definition 30 (Coalitional structure).** For a given coalitional game  $g = (A, \nu)$  the coalitional structure  $CS = \{C_1, C_2, \dots, C_m\}$  is the partition of the set  $A$ , which implies that  $\emptyset \notin CS$ ,  $\sum_{C \in CS} C = A$  and  $\forall_{C_1, C_2 \in CS} C_1 \neq C_2 \implies C_1 \cap C_2 = \emptyset$ .

One of the most important research topics in coalitional games is to find optimal coalitional structure [Sandholm et al., 1999, Rahwan et al., 2012]. More formally, this problem known as optimal coalition structure generation involves solving the following equation:  $\arg \max_{CS} \sum_{C \in CS} \nu(C)$ . Unfortunately, this problem is NP-hard [Sandholm et al., 1999].

Coalitional structure is a very similar to community structure defined on graphs (Definition 18). In this chapter we use this analogy and propose the first measure of node centrality that takes into account the *community structure* and that is based on coalitional games with coalitional structure.

### 6.2.2. The Owen Value

Coalitional games can be analysed from both the *ex ante* and *ex post* perspectives. In the *ex ante* perspective, it is not known which coalition will actually form. The Semivalues (Definition 7) are *ex ante* since they consider the sum of the marginal contributions of a player to all possible coalitions without any additional assumptions.

Another approach is to consider a coalitional game from the *ex post* perspective. In this case it is already known which coalitions form by the end of the game. In other words, it is known how the agents have partitioned themselves into a coalition structure. This *ex post* perspective is especially appealing for our model where networks are or can be divided into communities.

The most popular extension of the Shapley value (Definition 5) to “*ex post*”-like situations was proposed by Owen [1977]. To this end, let us first introduce the concept of the *quotient game*  $\nu^Q$ . Given the coalitional game  $(A, \nu)$  and coalitional structure  $CS = \{C_1, C_2, \dots, C_m\}$ , we define a new coalitional game, where the coalitions are considered to be individual players:

$$\nu^Q(R) = \nu\left(\bigcup_{r \in R} C_r\right) \text{ for all } R \subseteq M,$$

where the set  $M = \{1, 2, \dots, m\}$  represents coalitions’ numbers. Note that  $\bigcup_{r \in R} C_r$  is a *coalition of coalitions*. We will denote such coalition by  $Q_R$ .

We are now ready to define the Owen value. Given  $(A, \nu)$  and  $CS = \{C_1, C_2, \dots, C_m\}$ , the share of the grand coalition’s payoff,  $\nu(A)$ , given to player  $a_i \in C_j \in CS$  according to the Owen value is given by:

**Definition 31 (The Owen Value).** In a coalitional game  $(A, \nu)$  with coalition structure  $CS$  the Owen value for a player  $a_i$  is given by:

$$\phi_i^{OV}(A, \nu, CS) = \sum_{R \subseteq M \setminus \{j\}} \frac{1}{|M| \binom{|M|-1}{|R|}} \sum_{C \subseteq C_j \setminus \{a_i\}} \frac{1}{|C_j| \binom{|C_j|-1}{|C|}} MC(Q_R \cup C, i).$$

Let us examine the above formula more closely. The computation of the Owen value can be thought of as a two-step process. In the first step, coalitions play the game  $(M, \nu^Q)$  between themselves and receive their Shapley values. In the second step, the values of these communities are, in turn, divided among their members according to the Shapley value of the members.

The Owen value is the unique division scheme that satisfies the following five often desirable properties: Efficiency, Symmetry, Null player, Additivity, and Component Symmetry [Owen, 1977], which will be discussed more extensively in the next section.

### 6.3. The Coalitional Semivalue

In this chapter we introduce a generalization of the Owen value, where more general division schemes—Semivalues—are used as opposed to the Shapley value within the definition of Owen value. Specifically, we combine formula for Semivalue (Definition 7) with the formula for the Owen value (Definition 31) and propose *coalitional semivalues*.

**Definition 32 (The Coalitional Semivalue).** In a coalitional game  $(A, \nu)$  with coalition structure  $CS$  the Coalitional Semivalue for a player  $a_i$  is given by:

$$\phi_i^{CSEMI}(\nu, CS) = \sum_{0 \leq k < |M|} \beta(k) \sum_{0 \leq l < |C_j|} \alpha_j(l) \mathbb{E}_{T^k, C^l} [MC(Q_{T^k} \cup C^l, i)].$$

where  $T^k$  is a random set of size  $k$  drawn uniformly from the set  $M \setminus \{j\}$ , and  $C^l$  a the random set of size  $l$  drawn uniformly from the set  $C_j \setminus \{i\}$ . The function  $\beta : \{0, 1, \dots, |M| - 1\} \rightarrow [0, 1]$  is a function such that  $\sum_{k=0}^{|M|-1} \beta(k) = 1$ .  $\{\alpha_j\}_{j \in \{1, \dots, |M|\}}$  is a family of functions such that  $\alpha_j : \{0, 1, \dots, |C_j| - 1\} \rightarrow [0, 1]$  and  $\sum_{k=0}^{|C_j|-1} \alpha_j(k) = 1$ .

Intuitively,  $\beta$  is a probability distribution used to compute  $\phi_j(M, \nu^Q)$ , and  $\alpha_j$  is the probability distribution used to evaluate the players inside a coalition  $\phi_i(C_j, \nu)$ . Importantly, as shown in Table 6.1, by adopting various probability distributions, we can obtain the Owen value [Owen, 1977], as well as all of its modifications proposed to date in the literature: *Owen-Banzhaf value* [Owen, 1982], *symmetric coalitional Banzhaf value* [Amer et al., 2002], and *symmetric coalitional p-binomial semivalues* [Carreras and Puente, 2012].<sup>1</sup>

<sup>1</sup> We refrain from the axiomatic characterization of the new solution concept as being out of the scope of this dissertation.



Table 6.1: Values of  $\alpha$  and  $\beta$  for the Owen value and its various extensions.

Solution name	$\beta(k)$	$\alpha_j(l)$
Owen value [Owen, 1977]	$\frac{1}{ M }$	$\frac{1}{ C_j }$
Owen-Banzhaf value [Owen, 1982]	$\frac{\binom{ M -1}{k}}{2^{ M -1}}$	$\frac{\binom{ C_j -1}{l}}{2^{ C_j -1}}$
symmetric coalitional Banzhaf value [Amer et al., 2002]	$\frac{\binom{ M -1}{k}}{2^{ M -1}}$	$\frac{1}{ C_j }$
symmetric coalitional p-binomial semivalue [Carreras and Puente, 2012]	$p^k(1-p)^{ M -1-k}$ $p \in [0, 1]$	$\frac{1}{ C_j }$

## 6.4. The New Centrality Measure and Its Properties

Let us now introduce the game-theoretic network centrality measure based on coalitional semivalues:

**Definition 33 (The Coalitional Semivalue-based Centrality).** *The game-theoretic network centrality for the graph  $G$  with community structure  $CS$  is a quadruple  $(G, CS, \psi, \phi^{CSEMI})$ , where the value of each node  $v \in V$  is given by  $\phi_v^{CSEMI}(\nu_G, CS)$ , where  $\psi(G) = \nu_G$ .*

This is the first centrality measure that evaluates nodes by taking into account the community structure of the network. In the next section, we will consider various properties of this new measure.

The aim of the rest of this section is to translate the properties of various instances of coalitional semivalues into the properties of the resulting centrality measure. The first three properties are derived from Null Player, Additivity and Symmetry, respectively. Let  $(G, CS, \psi, \phi)$  be a game-theoretic network centrality for the graph  $G$  with community structure  $CS$  and  $v \in C_j \in CS$  some node from community  $C_j$ .

**Property 1.** *If a node makes no contribution to any community then its value is zero:*

$$\forall_{C \subseteq V \setminus \{v\}} \text{MC}(C, v) = 0 \implies \phi_v^{CSEMI}(\nu_G, CS) = 0.$$

**Property 2.** *If two group centralities are combined into one  $\nu_G = \nu'_G + \nu''_G$  then*

$$\phi_v^{CSEMI}(\nu_G, CS) = \phi_v^{CSEMI}(\nu'_G, CS) + \phi_v^{CSEMI}(\nu''_G, CS).$$

**Property 3.** *If two nodes from the same community  $v, u \in C_j$  contribute the same value to all possible communities then they are equally important:  $\forall_{C \subseteq V \setminus \{v, u\}} \text{MC}(C, v) = \text{MC}(C, u) \implies$*

$$\phi_v^{CSEMI}(\nu_G, CS) = \phi_u^{CSEMI}(\nu_G, CS).$$

The next property involves the Quotient game:

**Property 4.** *The power of a community is the aggregation of the power of nodes comprising this community. Formally:  $\phi_j^{SEMI}(M, \nu_G^Q) = \sum_{v \in C_j} \phi_v^{CSEMI}(\nu_G, CS)$ .*

Table 6.2: The properties of coalitional semivalue and its various instances.

Solution name	P1	P2	P3	P4	P5	P6
<b>coalitional semivalue [this chapter]</b>	✓	✓	✓	×	×	×
Owen value [Owen, 1977]	✓	✓	✓	✓	✓	✓
Owen-Banzhaf value [Owen, 1982]	✓	✓	✓	×	×	×
symmetric coalitional Banzhaf value [Amer et al., 2002]	✓	✓	✓	✓	×	✓
symmetric coalitionalp-binomial semivalue [Carreras and Puente, 2012]	✓	✓	✓	✓	×	✓

All the solutions, where the power inside the communities is computed using the Shapley value (due to the Efficiency), have the above property. In the same spirit, it can required that all the power of the whole network  $\nu_G(V)$  is distributed among the nodes:

**Property 5.** *The value of the whole network  $\nu(V)$  is the aggregation of the power of nodes comprising this network:  $\nu_G(V) = \sum_{v \in V} \phi_v^{CSEMI}(\nu_G, CS)$ .*

Our final property is the translation of Component Symmetry. If we define the marginal contribution of the coalition  $C$  to the set of nodes  $Q_T$  as  $\text{MC}(Q_T, C) = \nu(Q_T \cup C) - \nu(Q_T)$ , we get:

**Property 6.** *If two communities contribute the same value to all possible groups of communities then their evaluation is the same:  $\forall_{T \subseteq M \setminus \{i,j\}} \text{MC}(Q_T, C_i) = \text{MC}(Q_T, C_j) \implies \phi_i^{SEMI}(M, \nu_G^Q) = \phi_j^{SEMI}(M, \nu_G^Q)$ .*

Table 6.2 summarizes properties of the coalitional semivalues.

## 6.5. Computational Analysis

For many succinctly represented coalitional games, computing the Shapley value is NP-hard (in fact, it is often #P-complete (see Section 3.3 or [Chalkiadakis et al., 2011])). Naive algorithms to compute the Shapley value (exhaustively computing the average marginal contribution over all orderings of players) have exponential running time. Given this, there are two possible research directions. Firstly, efficient approximate algorithms can be developed. Secondly, classes of centralities, that have real-life applications and can be computed in polynomial time, can be defined. In this chapter, we take the latter approach and propose a polynomial time algorithm for computing coalitional semivalue-based centralities, where the characteristic function—the value of any group of nodes—is based on their degree. Thus, our method is build upon degree

centrality — an important method of evaluating nodes in social networks analysis [Freeman, 1979, Everett and Borgatti, 1999, Michalak et al., 2013a].

### 6.5.1. Weighted Degree Centrality Measure

In this subsection we define a class of cooperative games such that a node's value is based on its degree. The *weighted group degree centrality* of a community  $C$  in graph  $G$  is defined as follows:

$$\psi_{WD}(G) = \nu_G^{WD}(C) = \sum_{v \in N(C)} f(v), \quad (6.1)$$

where  $N(C)$  is the set of neighbours of  $C$ , and  $f$  is a parameter that is a polynomially computable function.

**Definition 34 (The Coalitional Semivalue-based Degree Centrality).** A *game-theoretic network weighted degree centrality* for the graph  $G$  with the community structure  $CS$  is a quadruple  $(G, CS, \psi_{WD}, \phi^{CSEMI})$ , where the value of each node  $v \in V$  is given by  $\phi_v^{CSEMI}(\nu_G^{WD}, CS)$ .

In the next section we will look more closely at the marginal contributions of nodes in order to effectively compute their expected value. This, in turn, will let us compute coalitional semivalues based on weighted group degree centrality in polynomial time.

### 6.5.2. The Marginal Contribution Analysis

In this subsection we lay the groundwork for the efficient algorithm that will compute coalitional semivalues for weighted group degree centrality in polynomial time. To this end, we will use equation from Definition 1. For a given node,  $v \in C_j \in CS$ , the focus will be on computing the expected value of its marginal contribution:  $\mathbb{E}_{T^k, C^l}[\text{MC}(Q_{T^k} \cup C^l, v)]$ .

We must consider the value of the expected marginal contribution of a node  $v$  to the set  $Q_{T^k} \cup C^l$ , where  $T^k$  is a random set of size  $k$ , and  $C^l$  is a random set of size  $l$ . Both sets are drawn uniformly from the sets  $M \setminus \{j\}$  and  $C_j \setminus \{v\}$ , respectively. We will construct the effective algorithm in two steps. First, we will decide under what conditions  $v$  makes a contribution to the set  $Q_{T^k} \cup C^l$ . Second, we will use a combinatorial argument to compute this contribution for each of the cases distinguished in the first step.

Before we start, we need to introduce the following notation: for the node  $v \in C_j \in CS$  we define the set of adjacent communities as  $N_{CS}(v) = \{C_i \in CS \setminus C_j \mid C_i \cap N(v) \neq \emptyset\}$ , inter-community degree as  $\deg_{CS}(v) = |N_{CS}(v)|$ , the set of neighbours within a community as  $N_j(v) = N(v) \cap C_j$ , and intra-community degree as  $\deg_j(v) = |N_j(v)|$ .

**Theorem 3.** The *game-theoretic network degree centrality* for graph  $G$  with community structure  $CS$  for the node  $v \in V$  can be computed in time polynomial in  $|V|$ .

*Proof.* In our proof we will use concepts from probability theory. Thus, we will first define the probability space, which is a triple  $(\Omega, \mathcal{F}, P)$ , where  $\Omega$  is a sample space containing sets  $\{Q_{T^k} \cup C^l\}$ ,  $T^k \subseteq 2^{|M|-1}$  is such that  $|T^k| = k$  and  $C^l \subseteq 2^{|C_j|-1}$  is such that  $|C^l| = l$ . The important observation is that  $|\Omega| = \binom{|M|-1}{k} \binom{|C_j|-1}{l}$ . In our model,  $\mathcal{F}$  is the set of elementary events ( $\mathcal{F} = \Omega$ ), and  $P : \mathcal{F} \rightarrow [0, 1]$  is a probability distribution function such that for each event  $A \in \mathcal{F}$  we have  $P(A) = \frac{1}{|\Omega|}$ .

There are two types of marginal contribution that a node can make. For the first, let us consider the marginal contribution of a single vertex  $v$  to the random set  $Q_{T^k} \cup C^l$ . When  $v$  joins a coalition  $C$ , it can contribute to its value with the help of any vertex  $u \in N(v)$  if and only if  $u$  is not in  $C$  and  $u$  is not already directly connected to  $C$ . Let us introduce the Bernoulli random variable  $B_{v,u,k,l}^{[1]}$ , which will indicate whether the vertex  $v$  makes a contribution through vertex  $u$  to the random set  $Q_{T^k} \cup C^l$ . Equation (6.1) tells us that this contribution will be  $f(u)$ . Thus, we have:

$$\mathbb{E}[f(u)B_{v,u,k,l}^{[1]}] = f(u)P[(N(u) \cup \{u\}) \cap (Q_{T^k} \cup C^l) = \emptyset],$$

where  $P[\cdot]$  denotes probability, and  $\mathbb{E}[\cdot]$  denotes expected value.

The second type of contribution takes place when vertex  $v$  joins a coalition  $C$  and takes away the value  $f(v)$ . Such a contribution happens when vertex  $v$  is directly connected to the coalition  $C$ . In particular, weighted group degree centrality  $\nu_G^D$  assumes that the value of a set of vertices depends only on nodes directly connected to this set, ignoring nodes already inside it. Therefore, when the node  $v$  becomes a member of  $C$  and it is not any more directly connected with it, the value of  $C$  is reduced by  $f(v)$ . Let us introduce the Bernoulli random variable  $B_{v,u,k,l}^{[2]}$ , which will indicate whether vertex  $v$  makes a contribution through itself to the random set  $Q_{T^k} \cup C^l$ . More formally, we have:

$$\mathbb{E}[-f(v)B_{v,k,l}^{[2]}] = -f(v)P[v \in N(Q_{T^k} \cup C^l)].$$

Now, we will move on to the second step of the proof and use a combinatorial argument to compute  $P^{[1]} = P[(N(u) \cup \{u\}) \cap (Q_{T^k} \cup C^l) = \emptyset]$  and  $P^{[2]} = P[v \in N(Q_{T^k} \cup C^l)]$ .

Recall that there are exactly  $\binom{|M|-1}{k} \binom{|C_j|-1}{l}$  sets  $Q_{T^k} \cup C^l$ . This is the size of the sample space. With this in mind, the probability  $P^{[1]}$  for  $u \in N(v)$ , if  $u, v \in C_j$  can be computed as follows:

$$P^{[1.1]} = \frac{\binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-1-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}},$$

otherwise if  $v \in C_j$  and  $u \in C_i$  and  $i \neq j$  we have:

$$P^{[1,2]} = \frac{\binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}}$$

Finally, for  $v \in C_j$  and  $u \in N(v)$  we obtain:

$$\mathbb{E}[f(u)B_{v,u,k,l}^{[1]}] = \begin{cases} 0 & \text{if } u \in C_j \\ & \text{and } (\deg_{CS}(u) > |M|-1 \\ & \text{or } \deg_j(u) > |C_j|-1) \\ f(u)P^{[1,1]} & \text{if } u \in C_j \\ 0 & \text{if } u \notin C_j \\ & \text{and } (\deg_{CS}(u) > |M|-1 \\ & \text{or } \deg_j(u) > |C_j|) \\ f(u)P^{[1,2]} & \text{if } u \notin C_j \end{cases} \quad (6.2)$$

In order to compute  $P^{[2]}$  we consider a complementary event  $P^{[2]} = (1 - P[N(v) \cap (Q_{T^k} \cup C^l)) = \emptyset])$  and using the same combinatorial argument as for computing  $P^{[1,1]}$ , for  $v \in C_j$  we get:

$$P^{[2]} = 1 - \frac{\binom{|M|-1-\deg_{CS}(v)}{k} \binom{|C_j|-1-\deg_j(v)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}},$$

and consequently we obtain:

$$\mathbb{E}[f(v)B_{v,k,l}^{[2]}] = \begin{cases} -f(v) & \text{if } (\deg_{CS}(u) > |M|-1 \\ & \text{or } \deg_j(u) > |C_j|-1) \\ -f(v)P^{[2]} & \text{otherwise.} \end{cases} \quad (6.3)$$

The final formula combines equations (6.2) and (6.3) :

$$\begin{aligned} \mathbb{E}[\text{MC}(Q_{T^k} \cup C^l, v)] &= \sum_{u \in N(v)} \left( \mathbb{E}[f(u)B_{v,u,k,l}^{[1]}] \right) + \mathbb{E}[f(v)B_{v,k,l}^{[2]}] \\ &= \sum_{u \in N(v) \cap C_j} f(u) \left( \frac{\binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-1-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}} \right) \\ &\quad + \sum_{u \in N(v) \setminus C_j} f(u) \left( \frac{\binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}} \right) \\ &\quad - f(v) \left( 1 - \frac{\binom{|M|-1-\deg_{CS}(v)}{k} \binom{|C_j|-1-\deg_j(v)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}} \right). \end{aligned} \quad (6.4)$$

The above formula can be used to compute  $\mathbb{E}_{T^k, C^l}[\text{MC}(Q_{T^k} \cup C^l, v)]$  in polynomial time. Therefore, the game-theoretic network degree centrality for graph  $G$  with community structure  $CS$  can be computed in polynomial time using equation from Definition 32 which ends our proof.  $\blacksquare$

### 6.5.3. Algorithms

Algorithm 17 directly implements expression from Definition 32. The expected value operator is computed using the final result of Theorem 3: equation (6.4). It computes the game-theoretic network degree centrality for a given graph  $G$  with community structure  $CS$ . For the sake of clarity, we assume in our algorithm that for  $a < b$  we have  $\binom{a}{b} = 0$ , and for any  $a$  we have  $\frac{a}{0} = 0$ .

---

**Algorithm 17:** The Coalitional Semivalue-based weighted degree centrality

---

**Input:** Graph  $G = (V, E)$ , node  $v \in V$ , coalition structure  $CS$ , functions  $\beta$  and family of functions  $\{\alpha\}$

**Data:** for each vertex  $u \in V$  and the community  $v \in C_j$ :

$\deg_{CS}(u)$  - the inter-community degree

$\deg_j(u)$  - the intra-community degree

**Output:**  $\phi_v^{CSEMI}$  coalitional semivalue-based degree centrality

```

1  $\phi_v^{CSEMI} \leftarrow 0$ ;
2 for  $k \leftarrow 0$  to  $|M| - 1$  do
3   for  $l \leftarrow 0$  to  $|C_j| - 1$  do
4      $\text{MC}_{k,l} \leftarrow 0$ ;
5     foreach  $u \in N(v) \cap C_j$  do
6        $\text{MC}_{k,l} \leftarrow \text{MC}_{k,l} + \frac{f(u) \binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-1-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}}$ 
7     foreach  $u \in N(v) \setminus C_j$  do
8        $\text{MC}_{k,l} \leftarrow \text{MC}_{k,l} + \frac{f(u) \binom{|M|-1-\deg_{CS}(u)}{k} \binom{|C_j|-\deg_j(u)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}};$ 
9      $\text{MC}_{k,l} \leftarrow \text{MC}_{k,l} - f(v)$ ;
10     $\text{MC}_{k,l} \leftarrow \text{MC}_{k,l} + \frac{f(v) \binom{|M|-1-\deg_{CS}(v)}{k} \binom{|C_j|-1-\deg_j(v)}{l}}{\binom{|M|-1}{k} \binom{|C_j|-1}{l}};$ 
11     $\phi_v^{CSEMI} \leftarrow \phi_v^{CSEMI} + \beta(k) \alpha_j(l) \text{MC}_{k,l};$ 
```

---

This algorithm requires some precomputations. For each node  $u \in V$  we need to calculate  $\deg_{CS}(v)$  and  $\deg_j(v)$ . We can store these values using  $O(|V|)$  space. Provided that it is possible to check the community of a given node in constant time, we can perform these precomputations in  $O(|V| + |E|)$  time. In the worst case, the main algorithm works in  $O(|V|^3)$  time.

Our next observation is that for trivial coalition structures (such as  $CS = \{A\}$ , or  $CS = \{\{a_1\}, \{a_2\}, \dots, \{a_n\}\}$ ) our algorithm computes any weighted degree-based semivalue in  $O(|V|^2)$  time. Finally, we would like to note that this algorithm is easily adapted to directed

---

**Algorithm 18:** The Owen value-based weighted degree centrality

---

**Input:** Graph  $G = (V, E)$ , node  $v \in V$ , coalition structure  $CS$ , functions  $\beta$  and family of functions  $\{\alpha\}$

**Data:** for each vertex  $u \in V$  and the community  $v \in C_j$ :

$deg_{CS}(u)$  - the inter-community degree

$deg_j(u)$  - the intra-community degree

**Output:**  $\phi_v^{CSEMI}$  coalitional semivalue-based degree centrality

```
1  $\phi_v^{CSEMI} \leftarrow 0$ ;  
2  $MC \leftarrow 0$ ;  
3 foreach  $u \in N(v) \cap C_j$  do  
4    $MC \leftarrow MC + \frac{f(u)}{(1+deg_{CS}(u))(1+deg_j(u))}$   
5 foreach  $u \in N(v) \setminus C_j$  do  
6    $MC \leftarrow MC + \frac{f(u)}{(1+deg_{CS}(u))deg_j(u)}$   
7  $MC \leftarrow MC + \frac{f(u)(1-(1+deg_j(v))(1+deg_{CS}(u)))}{(1+deg_j(v))(1+deg_{CS}(u))}$ ;  
8  $\phi_v^{CSEMI} \leftarrow \phi_v^{CSEMI} + MC$ ;
```

---

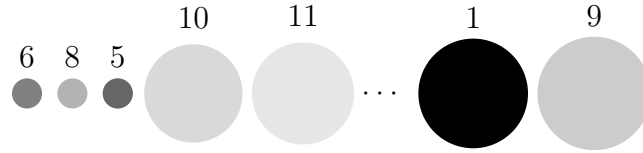


Figure 6.1: The relative power of communities for the first top nodes from the  $\nu_G^{WD}(\{v\})$  ranking. The power of the communities of nodes 5, 6 and 8 is significantly smaller than the power of communities of the other top nodes.

networks. To this end, depending on the new definition of weighted group degree, we need to replace all instances of  $deg_{CS}(u)$  and  $deg_j(u)$  with their counterparts for directed networks: *in* or *out* degree.

Algorithm 18 is the optimized version of Algorithm 17 to compute Owen value for weighted degree centrality.

## 6.6. Simulations

The main aim of this experiment is to compare three rankings created by three different methods: (i) one that uses weighted degree centrality and evaluates each node  $v$  by the number of neighbours it has (we denote it by  $\nu_G^{WD}(\{v\})$ ); (ii) one with the Shapley value-based degree centrality (denoted  $\phi_v^{SV}$ ); and, (iii) one with the Owen value-based degree centrality (denoted  $\phi_v^{OV}$ ), which evaluates nodes in the context of the communities they belong to and their respective power. Thus, the first two methods do not account for the existence of the community structure while the third one does.

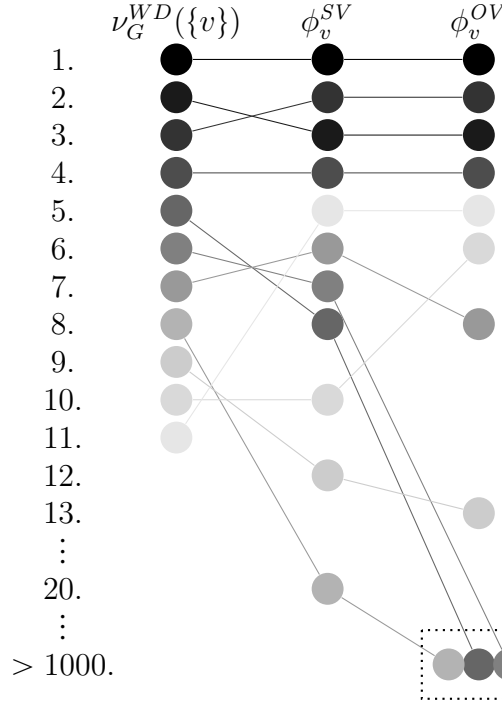


Figure 6.2: Three rankings of the top nodes. The *OV* ranking radically decreases the positions of the nodes 5, 6 and 8.

The real-life network used for simulations is a citation network that consists of 2,084,055 publications and 2,244,018 citation relationships.<sup>2</sup> This dataset is a list of publications with basic attributes (such as: title, authors, venue, or citations), and it is part of the project *ArnetMiner* being under development by Tang et al. [2008]. All publications extracted from this dataset were categorized into 22954 unique communities representing journals, conference proceedings or single book titles using basic text mining techniques. These communities can be interpreted as scientific groups united under the same topics of interests. In our experiment we use the directed version of our algorithm and assume that  $f(v) = \frac{1}{\text{\#number of articles citing } v}$ . The Shapley value-based centrality (the second method) is computed using the polynomial time algorithm introduced by Michalak et al. [2013a]. The Owen value-based centrality is computed with the modification of Algorithm 17, in which thanks to the form of the  $\alpha$  and  $\beta$  (in Owen value these discrete probabilities are uniform) the complexity was reduced to  $O(|V| + |E|)$ .

In what follows we focus on the 11 top nodes from the basic ranking  $\nu_G^{WD}(\{v\})$ . Figure 6.1 shows the relative power of the communities to which these nodes belong. Nodes indexed 5, 6 and 8 belong to significantly less powerful communities than nodes 1, 2, 3, 4, 7, 9, 10 and 11.

Figure 6.2 shows how the position of top nodes selected using  $\nu_G^{WD}(\{v\})$  changes in the  $\phi_v^{SV}$  and  $\phi_v^{OV}$  rankings. While for most nodes the perturbations are not so intensive, we observe

<sup>2</sup> The database used for these experiments is available under the following link: <http://arnetminer.org/citation>.



Table 6.3: The values of different coalitional semivalues.

<b>Solution name / Nodes's degree</b>	17	16	11	10	9
Owen value	3.51	2.68	1.47	1.37	0.70
Owen-Banzhaf value	2.28	1.38	0.47	0.88	0.01
symmetric coalitional Banzhaf value	3.51	2.68	1.47	1.37	0.70
symmetric coalitional p-binomial semivalue ( $p = \frac{1}{4}$ )	4.38	3.15	1.65	2.38	1.04

significant downgrade of the position of nodes 5, 6 and 8 in the  $OV_v$  ranking. This demonstrates coalitional semivalues-based centrality (in this case the Owen value-based centrality) is able to recognize that these three nodes belong to much weaker communities.

The rankings of nodes may differ depending which coalitional semivalue we choose. To illustrate this fact we evaluated top 5 nodes with the highest degree centrality from Zachary Karate Club Network [Zachary, 1977]. This network consists of 34 nodes divided into two communities. We observe in Table 6.3 that the ranking created with the Owen value differs with the one created with Owen-Banzhaf value at the 3<sup>rd</sup> and 4<sup>th</sup> positions.

## 6.7. Conclusions

The centrality metric proposed in this chapter is the first tool that evaluates individual nodes *in the context of their communities*. This metric is based on the Owen value—a well-known concept from coalitional game theory that we generalize by introducing coalitional semivalues. Our experiments show that the rankings can significantly differ if we account for the power of the relevant communities that the nodes belong to. If the community of a node is weak, it can significantly weaken the position of the node in the ranking based on the coalitional semivalue. It also demonstrates that our polynomial time algorithm is applicable to large data sets.

## 7. Computational Analysis of a Connectivity Game on Networks

We study a recently developed centrality metric to identify key players in terrorist organisations due to Lindelauf et al. [2013]. This metric, which involves computation of the Shapley value for *connectivity games on graphs* proposed by Amer and Gimenez [2004], was shown to produce substantially better results than previously used standard centralities. In this chapter, we present the first computational analysis of this class of coalitional games, and propose two algorithms for computing Lindelauf *et al.*'s centrality metric. Our first algorithm is exact, and runs in time linear by number of connected subgraphs in the network. As shown in the numerical simulations, our algorithm identifies key players in the WTC 9/11 terrorist network, constructed of 36 members and 125 links, in less than 40 minutes. In contrast, a general-purpose Shapley value algorithm would require weeks to solve this problem. Our second algorithm is approximate and can be successfully used to study much larger networks.

The contribution of the author of this dissertation covers the complexity analysis and developing exact and approximate algorithms for computing Lindelauf *et al.*'s centrality metric.

### 7.1. Introduction

Despite enormous efforts to tackle terrorism in the aftermath of 9/11, many terrorist networks are growing in size [Shultz et al., 2009]. For instance, the number of core members of Al'Qaeda, arguably the most infamous terrorist group, is currently estimated to exceed 800, with their operations spreading over more than 20 countries.

Facing, on the one hand, the increased size of terrorist groups and, on the other hand, inevitable budget cuts, security agencies urgently require efficient techniques to identify who plays the most important role within a terrorist network and, therefore, where scarce resources should predominantly be focused. In this context, a number of authors have proposed using *social network analysis* to investigate terrorist organisations (*e.g.*, Carley et al. [2003], Krebs [2002], Farley [2003], Lindelauf et al. [2013], Ressler [2006]).

Now, the key strength of social network analysis is its bottom-up approach in that the structure and the functioning of a group is gradually revealed by considering pairwise relationships between the individuals who create the group (*cfr.* Ressler [2006]). In this

context the identification of the key individuals within the network is called a *centrality analysis*. Analyst’s Notebook 8 [I2, 2010]—a software package used worldwide by law enforcement and intelligence agencies—has recently included standard centrality metrics for networks (*graphs*), such as degree, closeness and betweenness centralities [Brandes and Erlebach, 2005, Friedkin, 1991]. But the usefulness of these metrics for terrorist networks is limited as they are often unable to capture the complex nature of these organisations [Lindelauf et al., 2013].

Recently, to address these shortcomings, Lindelauf et al. [2013] developed a more advanced method specifically designed to measure centrality in terrorist networks. The new method belongs to a class of so-called *Shapley value-based centrality* (see Definition 25) and builds upon the notion of coalitional *connectivity games* proposed by Amer and Gimenez [2004]. These games are defined on graphs, where a coalition of nodes gets a non-zero value if and only if the induced sub-graph is connected. Such a game has an appealing interpretation in our context: it reflects communication capabilities among various groups of terrorists within the network. Indeed, when applied to the terrorist networks responsible for the 9/11 WTC and 2002 Bali attacks (with 23 and 17 members, respectively), Lindelauf *et al.* showed that their method produces qualitatively better results than standard centrality metrics.

Unfortunately, the computational aspects of connectivity games by Amer and Gimenez have not been studied to date. This means that the current use of Lindelauf *et al.*’s method is limited only to small terrorist networks (of 25 members or so) because general-purpose algorithms for coalitional games have to be applied, which exhaustively search the space of all possible coalitions. Thus, they are inapplicable to many real-world applications such as the terrorist networks responsible for the WTC 9/11 attack (from 35 to 63 nodes depending on the considered type of links between terrorist).

Against this background, we provide in this chapter the first computational analysis of connectivity games proposed by Amer and Gimenez:

- We prove that computing the Shapley value in connectivity games—including the centrality metrics of Lindelauf *et al.*—is #P-Hard.
- We propose a *dedicated exact algorithm* for computing these centrality metrics. While the general-purpose Shapley value algorithm requires checking all subsets of vertices in the graph, our algorithm traverses through (most often) much smaller number of connected subgraphs. It also has minimal memory requirements.
- We test our algorithm by analysing the aforementioned WTC 9/11 terrorist network with 36 members and 125 identified connections. In this setting, our algorithm returns the solution within 38 minutes, compared to weeks if a general-purpose approach was applied.
- In order to study even bigger networks, we propose a *dedicated approximate algorithm* based on Monte Carlo sampling. By comparing to our exact algorithm, we show that, after

a reasonable number of iterations, the approximate algorithm yields a very accurate ranking of top nodes based on the centrality of Lindelauf *et al.*

## 7.2. Related Work

A rapidly growing body of work is directed to the analysis of terrorist organisations using the methods of social network analysis. A very good introduction to this line of research can be found in Ressler [2006]. Also worthy of note is the work of Farley [2003], Carley *et al.* [2003], and Husslage *et al.* [2012], who conduct quantitative analysis of the terrorist networks.

Since the work of Grofman and Owen [1982], a number of game-theoretic centrality measures have been developed either to enrich the existing well-known centralities or as completely new ones (see Chapter 3). In the terrorist network context, Lindelauf *et al.* [2009a] and Lindelauf *et al.* [2009b] employed the game-theoretic approach to analyze covert networks.

The hardness result presented in this chapter is consistent with other studies of the complexity of the Shapley value in various settings. For instance, computing the Shapley value was shown to be #P-Complete for weighted majority games [Deng and Papadimitriou, 1994] and in minimum spanning tree games [Nagamochi *et al.*, 1997]. Aziz *et al.* [2009] obtained negative results for a related problem of computing the Shapley-Shubik power index for the spanning connectivity games that are based on undirected, unweighted multigraphs. Also, Bachrach *et al.* [2008] showed that the computation of the Banzhaf index for connectivity games, in which agents own vertices and control adjacent edges and aim to become connected to the certain set of primary edges, is #P-Complete. A comprehensive review of these issues, including some positive results for certain settings, can be found in [Chalkiadakis *et al.*, 2011].

Finally, it should be mentioned that other types of connectivity games have been considered in the literature. These include vertex connectivity games proposed by Bachrach *et al.* [2008] and the spanning connectivity games proposed by Aziz *et al.* [2009]. The common denominator of these games is that they are interested in maintaining the connectivity between certain set of nodes and they further study the ability of each node to affect the outcome of these games. In particular, the concepts like Banzhaf power index and Shapley-Shubik power index are utilized to measure the influence of nodes [Bachrach *et al.*, 2008, Aziz *et al.*, 2009].

In the already classic work, Krebs [2002] applied standard centrality measures to determine the key players in the 9/11 terrorist network. Memon *et al.* [2008] developed an algorithm based on the two well known centrality measures from social network analysis to automatically detect the hidden hierarchy in terrorist networks. Recently, Xu *et al.* [2009] explored the use of social network analysis methods to analyze terrorist networks mostly focusing on assigning roles to actors in the network. Most of the above mentioned work in the literature focus on identifying key players in the terrorist networks. A complementary approach is proposed by Henke [2009]

that combines aspects of traditional social network analysis with a novel multi-agent framework that describes how terrorist groups survive despite the aggressive counterterrorist operations.

### 7.3. Connectivity Games for Terrorist Networks

Terrorist networks have been recently modeled using a weighted graph,  $G$  (Definition 13), composed of individual terrorists and labeled edges [Carley et al., 2003, Krebs, 2002, Farley, 2003, Lindelauf et al., 2013, Ressler, 2006]. Based on available intelligence, an edge represents, for instance, a communication link between two terrorists, and the weight of the edge represents the frequency with which that link is used. Weights can be associated not only with edges but also with vertices. This, as argued by Lindelauf *et al.*, allows for modeling additional information that intelligence agencies gather on individuals within the network, such as access to weapons, financial means, previous experience, and participation in a terrorist training camp. We will denote the weight of vertex  $v_i \in V(G)$  as  $\gamma(i) \in \Gamma(G)$ .

**Definition 35 (Vertex-weighted graph).** *The vertex-weighted graph  $G = (V, E, \gamma)$  is a simple/directed/weighted graph, with a function  $\gamma(v) : V \rightarrow \mathbb{R}$  that evaluates each node. The value  $\gamma(v)$  is called the weight of a node  $v$ .*

To address the limitations of standard centrality metrics, Lindelauf *et al.* proposed a new metric that builds upon *connectivity games* by Amer and Gimenez [2004]. In these coalitional games on graphs, all subsets of set  $V(G)$  are considered to be possible coalitions of terrorists. Apart from the empty set, every single coalition is classified as belonging to either the set of *connected coalitions* (denoted  $\mathcal{C}(G)$ ) or *disconnected coalitions* (denoted  $\tilde{\mathcal{C}}(G)$ ). We say that  $C$  is connected if between any two nodes in  $C$  there exists at least one path of which all nodes belong to  $C$ . Otherwise  $C$  is disconnected. Importantly, any two terrorists in a connected coalition are able to communicate with each other (*via* a path), whereas in a disconnected coalition this is not the case. In many connected coalitions, from this point of view, some nodes play more important role than others, as their removal makes a coalition disconnected. We will call them *pivotal*.

**Definition 36 (Pivotal node).** *Given connected coalition  $C \in \mathcal{C}(G)$ , a node  $v_i \in C$  is **pivotal** to  $C$  iff  $C \setminus \{v_i\} \notin \mathcal{C}(G)$ .*

If such node is removed, the coalition becomes disconnected. In our context, communication between all terrorists within the subgroup becomes impossible.

**Definition 37 (Connectivity game).** *The connectivity game is a coalitional game defined on graph  $(V, \nu_C)$ , where the characteristic function is defined:*

$$\nu_C(C) = \begin{cases} 1 & \text{if } C \in \mathcal{C}(G) \\ 0 & \text{otherwise.} \end{cases}$$

In such a game defined by Amer and Gimenez [2004] every connected coalition is evaluated to 1, and disconnected coalitions evaluated to 0. Lindelauf *et al.* extend this definition by assuming that values of connected coalitions may depend on the network in a variety of ways, *i.e.*, they can be a function of adjacent edges or nodes, their weights, *etc.* More formally:

**Definition 38 (Lindelauf centrality).** *The Lindelauf centrality is a game-theoretic centrality  $(\psi_f, \phi^{SV})$ , where the representative function  $\psi_f$  is defined:*

$$\psi_f(G) = \nu_f(C) = \begin{cases} f(C, G) & \text{if } C \in \mathcal{C}(G) \\ 0 & \text{otherwise.} \end{cases}$$

The exact definition of  $f$  depends on the availability of information and analytical needs. For instance, to analyse the Jemaah Islamiyah network responsible for the 2002 Bali attack in Indonesia, Lindelauf *et al.* assume:

$$f(C) = \frac{|E(C)|}{\sum_{(v_i, v_j) \in E(C)} \lambda(v_i, v_j)}, \quad (7.1)$$

that is  $f$  equals the number of edges in the connected coalition  $C$  (denoted by  $E(C)$ ) divided by their weight.

Lindelauf *et al.* show that the centrality ranking based on the Shapley value of the connectivity game is much more effective than degree, closeness and betweenness centralities in exposing the key players in the Bali attack. In particular, Azahari bin Husin—the network bomb expert who was considered the “brain” behind the entire operation—is ranked low by standard centralities but according to Lindelauf *et al.* metric is among top five Bali terrorists. Similarly, Feri (Isa)—again ranked low by standard centralities—was, in fact, the suicide bomber. Lindelauf *et al.* ranked him third.

For big networks networks, the only feasible approach to compute Shapley value currently outlined in the literature is Monte-Carlo sampling (see Section 4.4.1). However, this method is not only inexact, but can be also very time-consuming. For instance, as shown in our simulations, for a weighted network of about 16,000 nodes and about 120,000 edges, the Monte Carlo approach has to iterate 300,000 times through the entire network to produce the

approximation of the Shapley value with a 40% error margin.<sup>1</sup> Exponentially more iterations are needed to further reduce this error margin.

## 7.4. Computational Analysis & Algorithms

In this section, we first discuss the complexity of computing the centrality metrics of Lindelauf *et al.*. We then present our exact and approximate algorithms.

### 7.4.1. Complexity

First we show that, even for the simplest definition of the connectivity game, where  $\forall C \in \mathcal{C}(G) f(C, G) = 1$ , computing the Shapley value in an efficient way is impossible. The main problem of interest is as follows:

**Definition 39 (#CG-SHAPLEY).** *Given a connectivity game on graph  $G$ , where  $\forall C \in \mathcal{C}(G) f(C, G) = 1$ , we are asked to compute the Shapley value for each node in this graph.*

In the first step, let us introduce the following problem:

**Definition 40 (#CONNECTED-SPANNING-SUB (#CSS)).** *Given a graph  $G$ , we are asked to compute the number of connected spanning subgraphs in  $G$ .<sup>2</sup>*

This problem is #P-Complete even for bipartite and planar graphs [Welsh, 1997]. We will use this hardness result to prove #P-Completeness of the following problem:

**Definition 41 (#CONNECTED-INDUCED-SUB (#CIS)).** *Given a graph  $G$ , we are asked to compute the number of connected induced subgraphs in  $G$ .*

**Theorem 4.** *#CONNECTED-INDUCED-SUB is #P-Complete.*

*Proof of Theorem 1* We note first that it is possible to check in polynomial time if a given subgraph is induced and connected. Thus, since a *witness* can be verified in polynomial time, this problem is in #P. Now, we will reduce a #CSS instance to #CIS. To this end, given a graph  $G = (V, E)$ , we will construct a transformed graph  $G'$  and show that determining the number of connected induced subgraphs in  $G'$  allows us to easily compute the number of connected spanning subgraphs in  $G$ .

<sup>1</sup> See Section 7.5 for the exact definition of the error margin.

<sup>2</sup> A connected subgraph  $F$  of the given graph,  $G = (V, E)$ , is called *spanning* if it contains all nodes from  $G$ , i.e.,  $V(F) = V(G)$ .

Our transformed graph  $G'$  is constructed by adding to each edge in  $G$  an extra node. Then, to each node from original graph  $G$  we attach a clique  $K_A$  with  $A$  nodes. This reduction is shown in Figure 7.1. More formally we define the following graph  $G'$ :

$$\begin{aligned} V(G') &= V(G) \cup \{v_i : v \in V(G) \wedge i \in \{1, \dots, A\}\} \cup \\ &\quad \{vu : (v, u) \in E(G)\} \\ E(G') &= \{(v, uv) : (u, v) \in E(G)\} \cup \\ &\quad \{(v, v_i) : v \in V(G) \wedge i \in \{1, \dots, A\}\} \cup \\ &\quad \{(v_i, v_j) : v \in V(G) \wedge i, j \in \{1, \dots, A\} \wedge i < j\} \end{aligned}$$

Now, we arbitrarily choose some connected induced subgraph  $F$  of  $G'$ . Either this subgraph intersects with the original set of vertices  $V(G)$ , or it does not. In the latter case, subgraph  $F$  is contained within the single copy of  $K_A$  and there are exactly  $|V(G)|(2^A - 1)$  such subgraphs. In the former case, we can define some *pseudograph*  $F'$ , which consists of the following sets of vertices and edges:<sup>3</sup>

$$\begin{aligned} V(F') &= V(F) \cap V(G) \\ E(F') &= \{(u, v) : uv \in V(F) \wedge (u, v) \in E(G)\} \end{aligned}$$

Note that since  $F$  is connected,  $F'$  also has to be connected. There are exactly  $2^{|V(F')|A}$  choices of  $F$  that can give us a particular pseudograph  $F'$ . Now, we can compute:

$$M = |V(G)|(2^A - 1) + \sum_{F'} 2^{|V(F')|A} \quad (7.2)$$

which denotes the number of induced connected subgraphs in  $G'$ .

The crucial observation here is that, if  $V(F') = V(G)$ , then  $F'$  is a connected spanning subgraph of  $G$ . This holds because  $F$  is an induced connected subgraph. Now, let  $N$  denote the number of spanning connected subgraphs in  $G$ . Then, we can rewrite (7.2) as:

$$M = |V(G)|(2^A - 1) + N(2^{|V(G)|A}) + \sum_{F': V(F') \neq V(G)} 2^{|V(F')|A} \quad (7.3)$$

In order to compute  $N$  we would like to bound the expression  $X = M - N(2^{|V(G)|A})$ . This expression is the number of induced connected subgraphs in  $G'$  that do not correspond to any

---

<sup>3</sup> We note that this tuple is not necessarily a properly defined graph, since it can contain some edge  $(u, v)$  and does not contain node  $v$ .



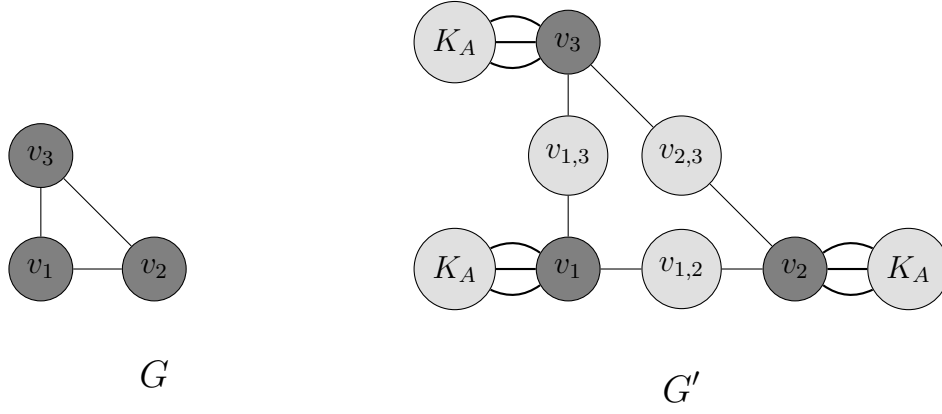


Figure 7.1: The reduction in the proof of Theorem 4.

connected spanning subgraph in  $G$ . We have:

$$0 \leq X \leq |V(G)|(2^A - 1) + 2^{|V(G)|+|E(G)|}2^{(|V(G)|-1)A}$$

where the number  $2^{|V(G)|+|E(G)|}$  is an upper bound on the number of all subgraphs in  $G$ . Now, we can use these bounds to transform equation (7.3) and to determine the approximation of  $N$ :

$$N = \frac{M}{2^{|V(G)|A}} - \frac{X}{2^{|V(G)|A}} \simeq \frac{M}{2^{|V(G)|A}} - \frac{|V(G)| + 2^{|V(G)|+|E(G)|}}{2^A}$$

In order to deal with the expression  $\frac{|V(G)|+2^{|V(G)|+|E(G)|}}{2^A}$  we take  $A > \log(|V(G)| + 2^{|V(G)|+|E(G)|})$  so that this fraction becomes smaller than 1. We note that  $A$  is bounded by a polynomial in the order of the input  $G$ . Then, the number of spanning connected subgraphs of  $G$  is the least integer  $N$  such that  $N \geq \frac{M}{2^{A|V(G)|}}$ . This is easy to compute given the number of connected induced subgraphs  $F$  of  $G'$ . This completes the proof.  $\square$

From Theorem 4, it trivially follows that the next problem is also #P-Complete:

**Definition 42 (#CONNECTED-INDUCED-SUB- $k$  (#CIS- $k$ )).** Given a graph  $G$ , we are asked to compute the number of connected induced subgraphs of size  $k$  in  $G$ .

Clearly, if we can find in polynomial time an answer for the #CIS- $k$  problem, we could efficiently compute #CIS. Now, the #P-Hardness of #CG-SHAPLEY will be shown by the reduction from #CIS- $k$ :

**Theorem 5.** #CG-SHAPLEY is #P-Hard.

*Proof of Theorem 2:* We construct a proof by reduction. In particular, we demonstrate that if there exists an algorithm for solving #CG-SHAPLEY in polynomial time, then it is possible to

solve #CIS- $k$  in polynomial time. This contradicts the fact that #CIS- $k$  is #P-complete. Now, we will reduce #CIS- $k$  to #CG-SHAPLEY.

Let  $G = (V, E)$  be an arbitrary graph, where  $|V| = n$  and  $|E| = m$ . We extend the set of nodes of  $G$  by a single node  $v$ , while the set of edges remains unchanged. In other words, we obtain a new graph  $G_0$  by adding a single node not connected to any node from  $G$ . Now, from the definition of connectivity games, we note that the marginal contribution of the new node  $v$  to any coalition  $C \subseteq G$  is either 0 or  $-1$ . More specifically, it is  $-1$  if  $C$  is connected, and 0 otherwise. Based on this, the Shapley value of node  $v$  can be computed as follows, where  $c_s^G$  is the number of connected induced subgraphs in  $G$  that contain exactly  $s$  nodes:

$$\phi_{v,G_0}^{SV} = -\sum_{s=0}^n \frac{(s)!(n-s)!}{(n+1)!} c_s^G$$

Now, let us consider a new graph  $G_i$  constructed by adding to  $G$  the set of  $i$  nodes in addition to the node  $v$ , while keeping the set of edges just as in  $G$ . Analogously to  $G_0$ , the Shapley value of  $v$  in  $G_i$  is:

$$\phi_{v,G_i}^{SV} = -\sum_{s=0}^n \frac{(s)!(n+i-s)!}{(n+1+i)!} c_s^G \quad (7.4)$$

This equation holds because each coalition containing more than  $n$  nodes is disconnected, and so the contribution of  $v$  to every such coalition is 0. Now, we can build a system of linear equations using each equation (7.4) from graph  $G_i$ , where  $i \in \{0, \dots, n\}$ . More precisely, we need to solve the following equation:

$$\begin{bmatrix} 0!n! & 1!(n-1)! & \cdots & n!0! \\ 0!(n+1)! & 1!n! & \cdots & n!1! \\ \vdots & \vdots & \ddots & \vdots \\ 0!(2n)! & 1!(2n-1)! & \cdots & n!n! \end{bmatrix} \begin{bmatrix} c_0^G \\ c_1^G \\ \vdots \\ c_n^G \end{bmatrix} = \begin{bmatrix} (n+1)!\phi_{v,G_0}^{SV} \\ (n+2)!\phi_{v,G_1}^{SV} \\ \vdots \\ (2n+1)!\phi_{v,G_n}^{SV} \end{bmatrix}$$

that can also be written as  $Ax = b$ .

This equation has a unique solution if and only if the determinant of the matrix  $A$  is non-zero. We can use Theorem 1.1 from [Bacher, 2002] to prove that it is non-zero. Thus, if we can compute in polynomial time the Shapley value for connectivity games, we would be able to solve this equation and determine all  $c_i^G$  values. Specifically, we can use Gaussian elimination, which works in  $O(n^3)$  time complexity.

We note that the largest possible number in our matrices is  $n!n!$ . According to the analysis in (Proposition 2)[Aziz et al., 2009] it is possible to store such a number in  $km^2(\log m)^2$  bits. It is shown in (Theorem 4.10) [Korte and Vygen, 2005] that in Gaussian elimination each number occurring during the algorithm process can be stored in the number of bits quadric of the input size. This final observation ends our proof.  $\square$

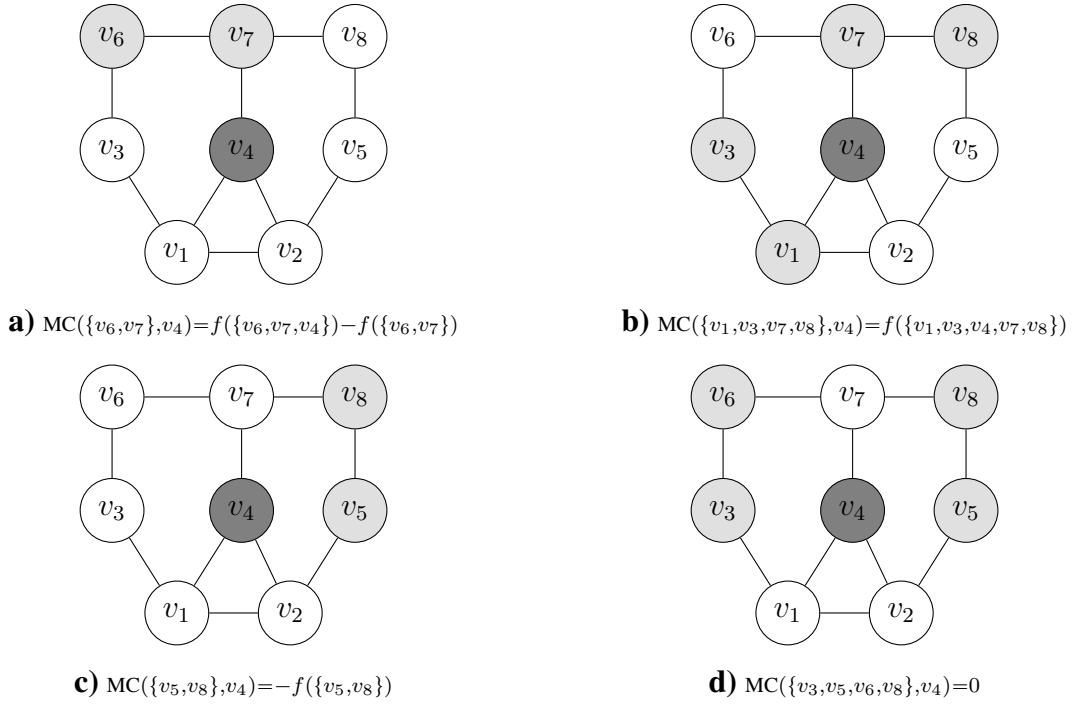


Figure 7.2: Four ways in which  $v_4$  can contribute to a coalition.

#### 7.4.2. Analysis of Marginal Contributions

In this section, we analyse how node  $v_i \in V$  can marginally contribute to coalition  $C \subseteq V \setminus \{v_i\}$ . Four general cases, depicted in Figure 7.2, can be distinguished:

- (a) Node  $v_i$  can join a *connected* coalition  $C \in \mathcal{C}$  and the resulting coalition is also *connected*, i.e.,  $C \cup \{v_i\} \in \mathcal{C}$ . Here, the marginal contribution is equal to the difference in the value of  $C \in \mathcal{C}$  caused by the addition of  $v_i$ :  

$$MC(C, v_i) = \nu_f(C \cup \{v_i\}) - \nu_f(C) = f(C \cup \{v_i\}) - f(C)$$
- (b) Node  $v_i$  can join a *disconnected* coalition  $C \in \tilde{\mathcal{C}}$  and the resulting coalition becomes *connected*, i.e.,  $C \cup \{v_i\} \in \mathcal{C}$ . Here,  $v_i$ 's contribution is the whole value of  $C \cup \{v_i\}$ :  

$$MC(C, v_i) = \nu_f(C \cup \{v_i\}) - \nu_f(C) = f(C \cup \{v_i\})$$
- (c) Node  $v_i$  can join a *connected* coalition  $C \in \tilde{\mathcal{C}}$  and the resulting coalition becomes *disconnected*, i.e.,  $C \cup \{v_i\} \in \tilde{\mathcal{C}}$ . This means that  $v_i$  brings down the value of  $C$  to 0:  

$$MC(C, v_i) = \nu_f(C \cup \{v_i\}) - \nu_f(C) = -f(C)$$
- (d) Node  $v_i$  can join a *disconnected* coalition  $C \in \tilde{\mathcal{C}}$  and the resulting coalition remains *disconnected*, i.e.,  $C \cup \{v_i\} \in \tilde{\mathcal{C}}$ :  

$$MC(C, v_i) = \nu_f(C \cup \{v_i\}) - \nu_f(C) = 0$$

The key conclusion to be drawn from the above analysis is that both connected and disconnected coalitions play a role when computing the Shapley value. This is because a node

can contribute not only to a connected coalition, but also to a disconnected one (by making it connected). However, in the further sections, we will show that it is possible to develop an exact algorithm that only cycles through connected coalitions.

### 7.4.3. The Basic Algorithm for Connectivity Games

Based on the observation that both connected and disconnected coalitions play a role when computing the Shapley value, we develop the general-purpose Shapley value algorithm (**GeneralSV** presented in Algorithm 19) for computing the Shapley value in connectivity games defined by Amer and Gimenez. In more details, for each coalition  $C \in 2^V$  this algorithm considers four aforementioned cases (a), (b) and (c) as outlined in Figure 7.2 in our chapter. Importantly, for (d), it can be disregarded since the marginal contribution in this case equals 0. In all of the rest three cases, a node  $v_i$  can make a non-zero contribution by joining a coalition. We note that function  $CheckConnectedness(C)$  runs in  $O(|C| + |E(C)|)$ .

---

#### Algorithm 19: GeneralSV Algorithm for the Shapley value

---

**Input:** Graph  $G = (V, E)$  and characteristic function  $\nu_f$   
**Output:** Shapley Value,  $\phi_i^{SV}(\nu_f)$ , for each node  $v_i \in V$

```

1 foreach  $v_i \in V$  do
2    $\phi_i^{SV}(\nu_f) \leftarrow 0$ ;
3 foreach  $C \in 2^V$  do
4    $CheckConnectedness(C)$ ;
5   if  $C \in \mathcal{C}$  then
6     foreach  $v_i \in N(C)$  do
7        $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|C|!(|V|-|C|-1)!}{|V|!}(\nu_f(C \cup \{v_i\}) - \nu_f(C))$ 
8     foreach  $v_i \notin N(C)$  do
9        $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) - \frac{|C|!(|V|-|C|-1)!}{|V|!}\nu_f(C)$ 
10  else
11    foreach  $v_i \in N(C)$  do
12       $CheckConnectedness(C \cup \{v_i\})$ 
13      if  $C \cup \{v_i\} \in \mathcal{C}$  then
14         $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|C|!(|V|-|C|-1)!}{|V|!}\nu_f(C \cup \{v_i\})$ 

```

---

Unfortunately, the use of this algorithm in practice is limited by the number of nodes in the network. It runs in  $O((|V| + |E|)2^{|V|})$  and already for  $|V| = 50$ , the Algorithm 19 has to cycle through more than  $10^{15}$  coalitions. In this dissertation, we developed an algorithm that for many networks is able to compute the SV substantially faster.

#### 7.4.4. The Faster Algorithm for Connectivity Games

Many real-world terrorist networks are sparse, *i.e.*,  $|\mathcal{C}| \ll |\tilde{\mathcal{C}}|$  [Krebs, 2002]; thus, if the Shapley value could be computed only by considering coalitions in  $\mathcal{C}$ , it would be possible to analyse much larger terrorist networks. To this end, for each  $v_i \in V$ , let us define the following disjoint sets of coalitions:

$$\begin{aligned}\mathcal{C}_i^\# &= \{C \subseteq V \setminus \{v_i\} : C \in \mathcal{C} \wedge C \cup \{v_i\} \in \mathcal{C}\} \\ \mathcal{C}_i^+ &= \{C \subseteq V \setminus \{v_i\} : C \in \tilde{\mathcal{C}} \wedge C \cup \{v_i\} \in \mathcal{C}\} \\ \mathcal{C}_i^- &= \{C \subseteq V \setminus \{v_i\} : C \in \mathcal{C} \wedge C \cup \{v_i\} \in \tilde{\mathcal{C}}\}\end{aligned}$$

which correspond to cases (a), (b) and (c) from the previous section. Based on this, the original formula for Shapley value (Definition 5) can be computed as:

$$\begin{aligned}\phi_i^{SV}(V, \nu_f) &= \sum_{C \subseteq V \setminus \{v_i\}} \frac{|C|!(|V| - |C| - 1)!}{|V|!} (\nu_f(C \cup \{v_i\}) - \nu_f(C)) \\ &= \sum_{C \in \{\mathcal{C}_i^+ \cup \mathcal{C}_i^\# \cup \mathcal{C}_i^-\}} \frac{|C|!(|V| - |C| - 1)!}{|V|!} (\nu_f(C \cup \{v_i\}) - \nu_f(C))\end{aligned} \quad (7.5)$$

where case (d), when  $mc_i(C) = 0$ , is simply omitted. The key idea behind our exact algorithm to compute the Shapley value in connectivity game (**FasterSVCG** presented in Algorithm 20) is to represent the sets  $\mathcal{C}_i^+$  and  $\mathcal{C}_i^-$  differently, such that  $\tilde{\mathcal{C}}$  does not appear in the new representation. As for  $\mathcal{C}_i^\#$ , it does not depend on  $\tilde{\mathcal{C}}$ , and so there is no need to represent it differently. In particular, we represent  $\mathcal{C}_i^+$  and  $\mathcal{C}_i^-$  as follows, where  $\mathcal{P}(C)$  is the set of agents that are pivotal to  $C$ , and  $N(C)$  is the set of neighbours of  $C$ :

$$\begin{aligned}\mathcal{C}_i^+ &= \{C \subseteq V \setminus \{v_i\} : C \cup \{v_i\} \in \mathcal{C} \wedge v_i \in \mathcal{P}(C \cup \{v_i\})\} \\ \mathcal{C}_i^- &= \{C \subseteq V \setminus \{v_i\} : C \in \mathcal{C} \wedge v_i \notin N(C)\}\end{aligned}$$

Now since  $\tilde{\mathcal{C}}$  no longer appears in the definitions of  $\mathcal{C}_i^\#$ ,  $\mathcal{C}_i^+$  and  $\mathcal{C}_i^-$ , it is possible to compute the Shapley value as in equation (7.5) *without enumerating any of the coalitions in  $\tilde{\mathcal{C}}$* . Based on this, our algorithm enumerates every connected coalition,  $C \in \mathcal{C}$ , and determines for each

agent  $v_i \in C$  whether  $C \setminus \{v_i\} \in \mathcal{C}_i^\#$  or  $C \setminus \{v_i\} \in \mathcal{C}_i^+$  and for  $v_i \notin C$  if  $C \in \mathcal{C}_i^-$ . Note, that we do not consider the impact of agent  $v_i \notin C$  that do not disconnect  $C$  because the contribution of this agent will be calculated for connected coalition  $C \cup \{v_i\}$  as  $C \in \mathcal{C}_i^\#$ . The enumeration is carried out using Moerkotte and Neumann [2006]’s method—the fastest such enumeration method in the literature. Its basic idea is that, for each connected coalition  $C \in \mathcal{C}$ , it expands  $C$  by adding to it certain subsets of its neighbours. These subsets are chosen so as to ensure that no connected coalition is enumerated more than once (see Moerkotte and Neumann [2006] for more details).

Next, we explain our algorithm. To enhance clarity, for every connected coalition  $C \in \mathcal{C}$ , we will define three disjoint sets of agents:  $V_C^\# = \{v_i \in C : C \setminus \{v_i\} \in \mathcal{C}_i^\#\}$ ,  $V_C^+ = \{v_i \in C : C \setminus \{v_i\} \in \mathcal{C}_i^+\}$ , and  $V_C^- = \{v_i \in V \setminus C : C \in \mathcal{C}_i^-\}$ . If we compute the above sets for every  $C \in \mathcal{C}$ , then we can compute the Shapley value. Let us take a closer look at the difficulty of computing those sets for a given  $C$ .

- Computing  $V_C^-$  can be done in  $O(|V|)$  time. This is because the agents in  $V_C^-$  are basically all those that are not members, nor neighbours, of  $C$ .
- Now, to compute  $V_C^+$ , we need to find the pivotal agents in  $C$ . This can be computed using a method “*findPivotals*” that runs in  $O(|V| + |E|)$  [Alsuwaiyel, 1999].
- Having computed  $V_C^+$ , it becomes easy to compute  $V_C^\#$ . This is because  $V_C^\# = C \setminus V_C^+$ .

From the above analysis, it is clear that the main difficulty lies in *findPivotals*. Therefore, whenever possible, we would like to compute  $V_C^+$  using some other, easier, technique. In particular, when we expand a connected coalition,  $C$ , into another connected coalition  $C' = C \cup S$ , we try to *update* the set of pivotal agents, rather than compute it from scratch with *findPivotals*. Here, we distinguish between three conditions:

- Condition 1: The cycles in  $C'$  are exactly like those in  $C$ . In this case, the set  $V_{C'}^+$  consists of elements of  $V_C^+$ , expanded by the nodes in  $C$  that are connected to  $S$ .
- Condition 2:  $C'$  contains a cycle that is not in  $C$ . Here, we need to call *findPivotals*.
- Condition 3:  $|C| = 2$ . In this case, since we assumed that a singleton is a connected coalition, none of the two agents in  $C$  is pivotal.

The pseudocode of FasterSVCG is presented in Algorithm 20. It is easy to see that it runs in  $O((|V| + |E|)|\mathcal{C}|)$ .

#### 7.4.5. Approximation algorithm for connectivity games

Our FasterSVCG algorithm is much faster than the GeneralSV, but it is not fast enough to deal with large networks. In order to provide ranking for larger networks we should use an approximation algorithms. In this section we provide the new approximation algorithm for

---

**Algorithm 20: FasterSVCG** Faster algorithm for the Shapley value
 

---

**Input:** Graph  $G=(V, E)$  and characteristic function  $\nu_f$   
**Output:** Shapley value  $\phi_i^{SV}(\nu_f)$  of each node  $v_i \in V$

```

1  $X \leftarrow V$ ; // initialize  $X$ , which is only used for Moerkotte & Neumann's enumeration
2 foreach  $v_i \in V$  do  $\phi_i^{SV}(\nu_f) \leftarrow 0$ ;
3 for  $i \leftarrow |V|$  to 1 do
4    $\text{computeSV}(\{v_i\}, N(v_i), X, X \setminus (N(v_i) \cup \{v_i\}), \emptyset)$ ;  $X \leftarrow X \setminus \{v_i\}$ 
   // ----- Next, we define computeSV -----
5 computeSV( $C, NC, X, V_C^-, V_C^+$ ) begin
6    $X' \leftarrow X \cup NC$ ; // where  $NC$  consists of the neighbours of  $C$ 
7   foreach  $S \subseteq (NC \setminus X) \wedge S \neq \emptyset$  do
8      $C' \leftarrow C$ ; // a new coalition that will be constructed from the old coalition  $C$ 
9      $NC' \leftarrow NC \setminus S$ ; // the neighbours of  $C'$ 
10     $isCycle \leftarrow false$ ; // to indicate whether a new cycle has appeared while
11    constructing  $C'$ 
12    foreach  $v \in S$  do
13       $TEMP \leftarrow \emptyset$ ; // a temporary set used to compute neighbours of:  $C' \cup \{v\}$ 
14       $TEMP2 \leftarrow \emptyset$ ; // a set used to compute the pivotal agents in:  $C' \cup \{v\}$ 
15      foreach  $u \in N(S)$  do
16        if  $u \notin (C \cup NC)$  then
17           $TEMP \leftarrow TEMP \cup \{u\}$ ;
18        else if  $(isCycle = false) \wedge (u \in C')$  then // condition 1
19           $TEMP2 \leftarrow TEMP2 \cup \{u\}$ ;
20      if  $|TEMP2| > 1$  then // condition 2
21         $isCycle \leftarrow true$ ;
22       $C' \leftarrow C' \cup \{v\}$ ;  $NC' \leftarrow NC' \cup TEMP$ ;
23       $V_{C'}^- \leftarrow V_{C' \setminus \{v\}}^- \setminus TEMP$ ;
24      if  $isCycle = false$  then // condition 1
25         $V_{C'}^+ \leftarrow V_{C' \setminus \{v\}}^+ \cup TEMP2$ ;
26      if  $|C'| = 2$  then // condition 3
27         $V_{C'}^+ \leftarrow \emptyset$ ;
28      else if  $isCycle = true$  then // condition 2
29         $V_{C'}^+ \leftarrow FindPivotals(C')$ ;
30      foreach  $v_i \in C'$  do // update Shapley value
31        if  $v_i \in V_{C'}^+$  then
32           $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|C' \setminus \{v_i\}|! (|V| - |C' \setminus \{v_i\}| - 1)!}{|V|!} \nu_f(C')$ 
33        else // deal with the set  $V_{C'}^\#$ 
34           $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|C' \setminus \{v_i\}|! (|V| - |C' \setminus \{v_i\}| - 1)!}{|V|!} (\nu_f(C') - \nu_f(C' \setminus \{v_i\}))$ 
35      foreach  $v_i \in V_{C'}^-$  do // update Shapley value
36         $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) - \frac{|C'|! (|V| - |C'| - 1)!}{|V|!} \nu_f(C')$ 
37     $\text{computeSV}(C', NC', X', V_{C'}^-, V_{C'}^+)$ ;

```

---

connectivity games (called **ApproximateSVCG**) that outperforms the standard Castro et al. [2009] method. To see discussion about different methods of approximation the Shapley value see Section 4.4.1.

Our Algorithm 21 is the dedicated application of standard Monte Carlo sampling to connectivity games. The crucial difference between Castro et al. [2009] method and ApproximateSVCG is that instead of sampling permutations our algorithm sample coalitions. The first observation is that during sampling a single set  $C$  we can compute two contributions: one  $\nu_f(C \cup v_i) - \nu_f(C)$ , and the second  $\nu_f(C) - \nu_f(C \setminus \{v_i\})$ . Generally speaking, in our algorithm, we will randomly select  $C$  and approximate the  $\phi_i^{SV}(\nu_f)$  using the resulting average. Looking at the equation (7.5) we see that the marginal contributions are calculated with different weights. Thus, to obtain an unbiased estimator we have to compute marginal contributions with appropriate probabilities. To this end, we propose the following process consisting of 3 fazes. In **Faze 1**, we uniformly select the size of the coalition  $k \in \{0, \dots, |V|\}$ . In **Faze 2**, we choose a random coalition  $C$  of size  $k$  and, in **Faze 3**, for every agent, compute the marginal contribution of this agent obtained by leaving/entering  $C$ . To better explain our motivation, let us transform the formula for the Shapley value (Definition 5) as follows:

$$\phi_i^{SV}(\nu_f) = \frac{1}{|V|} \sum_{0 \leq k < |V|} \cdot \frac{1}{\binom{|V|-1}{k}} \sum_{C \subseteq V \setminus \{i\}, |C|=k} (v(C \cup \{v_i\}) - v(C)) \quad (7.6)$$

The similar transformation was already made in equation 5.1, when we used it to compute efficiently Semivalues.

From the above formula we infer that to obtain an unbiased estimator the sampling method should satisfy two conditions:

- (i) the probability that a marginal contribution to randomly chosen  $C$  is obtained from entering  $v_i$  to a coalition of size  $k$  is equal for every  $k$ :  $\frac{1}{n+1} \cdot \frac{n-k}{n} + \frac{1}{n+1} \cdot \frac{k+1}{n} = \frac{1}{n}$  ( $n = |V|$ );<sup>4</sup>
- (ii) marginal contributions to all coalitions of size  $k$  are chosen with the same probability.

For instance consider a pivotal node  $v_i$  in coalition  $C$  of size  $k + 1$ . The expected marginal contribution when  $C \in \tilde{\mathcal{C}}$  and  $C \cup \{v_i\} \in \mathcal{C}$  equals  $\frac{1}{n+1} \cdot \binom{1}{\binom{n-1}{k-1}} \cdot \nu_f(C)$ . On the other hand, the same marginal contribution is obtained from leaving connected coalition  $C$  – here expected value equals  $\frac{1}{n+1} \cdot \binom{1}{\binom{n-1}{k}} \cdot \nu_f(C)$ . As we won't consider transfers from the disconnected coalitions, we have to ensure that expected value from this marginal contribution does not change, so if we obtain  $C$  with the probability  $\frac{1}{n+1} \cdot \binom{1}{\binom{n-1}{k}}$  we add marginal contribution which equals  $\nu_f(C) \cdot (1 + \frac{n-k+1}{k}) = \nu_f(C) \cdot \binom{n+1}{k}$ .

---

<sup>4</sup> Note that a given marginal contribution appears twice in our process—the first term represents the probability that we select a coalition of size  $k$  without player  $v_i$ , while the second one—that we select the corresponding coalition of size  $k + 1$ , with player  $v_i$ .



---

**Algorithm 21: ApproximateSVCG** Approximation algorithm for the Shapley value

---

**Input:** Graph  $G=(V, E)$  and characteristic function  $\nu_f$   
**Output:** Shapley value  $\phi_i^{SV}(\nu_f)$  of each node  $v_i \in V$

```
1 foreach  $v_i \in V$  do  $\phi_i^{SV}(\nu_f) \leftarrow 0$ ;  
2 for  $it = 1$  to  $maxIter$  do  
3    $k \leftarrow \text{random number from } \{0, \dots, |V|\}$ ;  
4    $C \leftarrow \text{random coalition of size } k$ ;  
5   if  $!CheckConnectedness(C)$  then continue;  
6    $P \leftarrow FindPivotal(C)$ ;  
7   foreach  $v_i \in C \setminus P$  do // case (a)  
8      $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|V|+1}{|C|} \cdot (\nu_f(C) - \nu_f(C \setminus \{v_i\}))$   
9   foreach  $v_i \in P$  do // case (b)  
10     $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) + \frac{|V|+1}{|C|} \cdot \nu_f(C)$   
11   foreach  $v_i \in (V \setminus C) \setminus N(C)$  do // case (c)  
12     $\phi_i^{SV}(\nu_f) \leftarrow \phi_i^{SV}(\nu_f) - \frac{|V|+1}{|V|-|C|} \cdot \nu_f(C)$   
13 foreach  $v_i \in V$  do  $SV_i(\nu_f) \leftarrow SV_i(\nu_f)/maxIter$ ;
```

---

This technique allows us to compute the marginal contributions of all agents for a randomly selected coalition  $C$ , which in the connectivity games can be performed much faster than estimating the Shapley value for each player separately [Mann and Shapley, 1962] or by sampling of a random permutation [Castro et al., 2009], where we have to calculate the marginal contributions for a sequence of coalitions growing in size.

In the ApproximateSVCG algorithm we merge our technique with the analysis of marginal contribution presented in Figure 7.2. The pseudo code is presented in Algorithm 21. Line 3 corresponds to **Faze 1**, where we sample the size of a coalition. Now, we modify **Faze 2** in order to select only *connected coalitions* (line 5). See that generating a random *connected coalition* uniformly will create a biased algorithm. We also modify **Faze 3**, when we consider cases (a), (b), and (c) from Figure 7.2 (lines 7-12). The modification of **Faze 3** has to be done due to the following reason: since we no longer consider disconnected coalitions, any non-zero marginal contribution made to a disconnected coalition have to be transfered to a corresponding connected coalition (lines 8, 10, and 12). Furthermore, this should be done in a way that preserves appropriate probabilities (thus, in lines 8, 10, and 12 we multiply marginal contributions by adequate weights). Finally, we divide the sum of the contributions by the number of iterations (lines 13). For every sample, algorithm runs in time  $O(|V| + |E|)$ .

## 7.5. Performance Evaluation

In this section, we empirically evaluate our FasterSVCG algorithm focusing on randomly generated networks as well as on real-life data from 9/11 terrorists network. We also evaluate

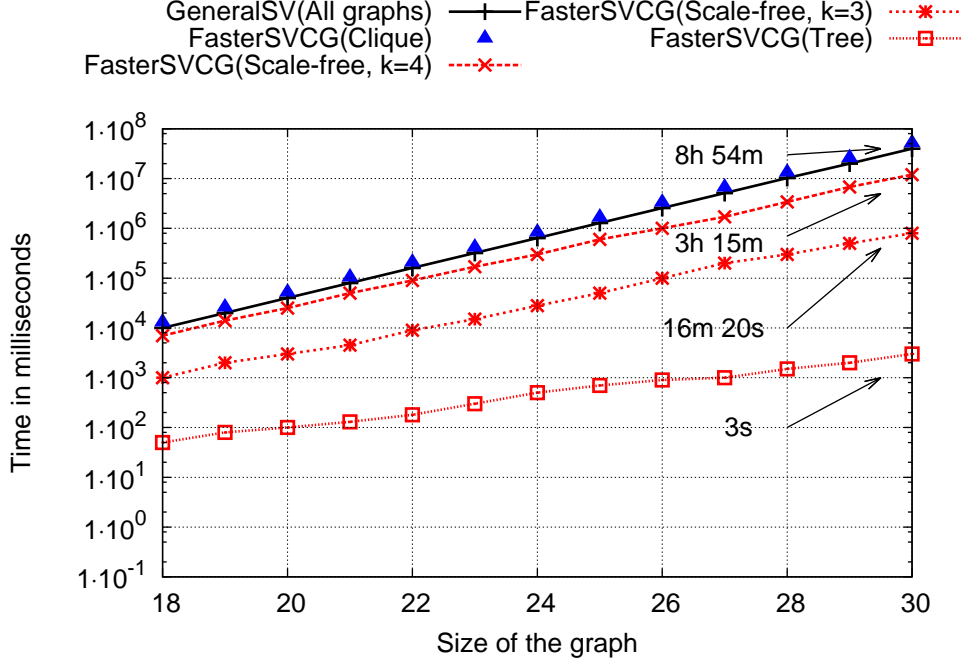


Figure 7.3: Time performance of both algorithms.

how ApproximateSVCG approximates ranking, by benchmarking it against exact solution obtained by our algorithm.

The random graphs used in experiments are based on two topologies commonly found in social networks, and in terrorist organizations in particular [Krebs, 2002, Sageman, 2004]: (i) *scale-free graphs*, where the network is generated according to a power law; and (ii) *random trees*, which model hierarchical organisations. To construct scale-free graphs, we use the preferential attachment generation model [Albert et al., 2000], with parameters  $k = 2, 3$ . In this model, while gradually constructing a graph, every new node  $v_i$  is linked to  $k$  already existing nodes such that the probability that  $v_i$  is linked to the node  $v_j$  is  $\frac{\text{degree}(v_j)}{\sum_j \text{degree}(v_j)}$ . To construct trees, every new node is attached to a randomly picked incumbent. Finally, we include in our analysis (iii) *complete graphs*, where all coalitions are connected. Although complete graphs are unlikely to arise in a terrorist network context, they constitute a suitable benchmark for our simulations. For all the games on random graphs we assume that  $f$  is defined as in Lindelauf's centrality (Definition 38, equation 7.1).

**FasterSVCG:** In Figure 7.3, we compare the performance of FasterSVCG to the general-purpose Shapley value algorithm (called GeneralSV and presented in Algorithm 19). This latter algorithm does not take into account the structure of the network, so its runtime is the same for any type of graph. Unlike GeneralSV, FasterSVCG takes advantage of the sparsity of the network and, consequently, significantly outperforms the benchmark. For instance, for SFG( $k = 2$ ) and  $|V| = 30$ , FasterSVCG needs only 0.39% of GeneralSV runtime! Naturally,

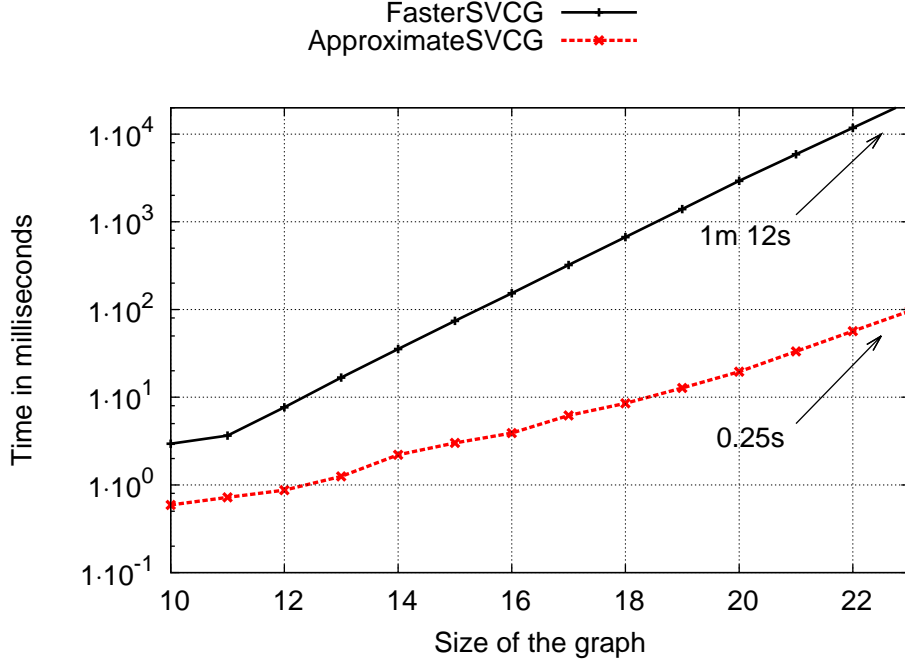


Figure 7.4: Time performance of our exact and approximation algorithms.

the best performance is obtained for tree graphs, where the game over a network of 30 nodes can be computed in about 3 seconds.

**ApproximateSVCG:** In Figure 7.4, we evaluate the time performance of ApproximateSVCG. Importantly, for our purpose of identifying key terrorists, we are mostly interested in the approximation of the correct Shapley value-based ranking of top nodes, and less in the approximation of their actual Shapley values. To this end, we evaluate the time required by ApproximateSVCG to obtain the ranking of the top  $\lceil \sqrt{n} \rceil$  nodes, with at most one error (i.e., one inversion compared to the exact ranking). We present an average time calculated over 500 iterations, each for a randomly selected SFG( $k = 4$ ). ApproximateSVCG returns top nodes much faster than FasterSVCG—for the graph of 24 nodes the former algorithm runs in 0.25 sec. and the exact one in 72 sec.

Figure 7.5 shows the comparison between the error convergence of the ApproximateSVCG and the error convergence of random permutation sampling studied by Castro et al. [2009] (see Algorithm 6). In this experiment we focus on the maximum absolute error of the Shapley value. It is computed as a percentage of the value of the grand coalition. The plot presents an average error from over 30 iteration for Krebs' 9/11 WTC terrorist network with 36 nodes. Since connected coalitions constitute only 0.59% of all coalitions, ApproximateSVCG outperforms Castro's method. Indeed, after 4 seconds, the average error of ApproximateSVCG equals 0.029%, while the error for Castro et al. exceeds 0.2%. The absolute error in ApproximateSVCG converges to zero which indicates that this sampling method is not biased.

	Lindelauf et al. $ V  = 19,  E  = 32$	Our analysis $ V  = 26,  E  = 125$
1.	A. Aziz Al-Omari	Z. Moussaoui
2.	H. Alghamdi	A. Aziz Al-Omari
3.	W. Alsheri	M. Atta
4.	H. Hanjour	W. Alshehri
5.	M. Al-Shehhi	N. Alhazmi

Table 7.1: FasterSVCG allows us to consider the bigger network of WTC 9/11 attack than Lindelauf *et al.* which delivers new insights into the leadership structure of this network

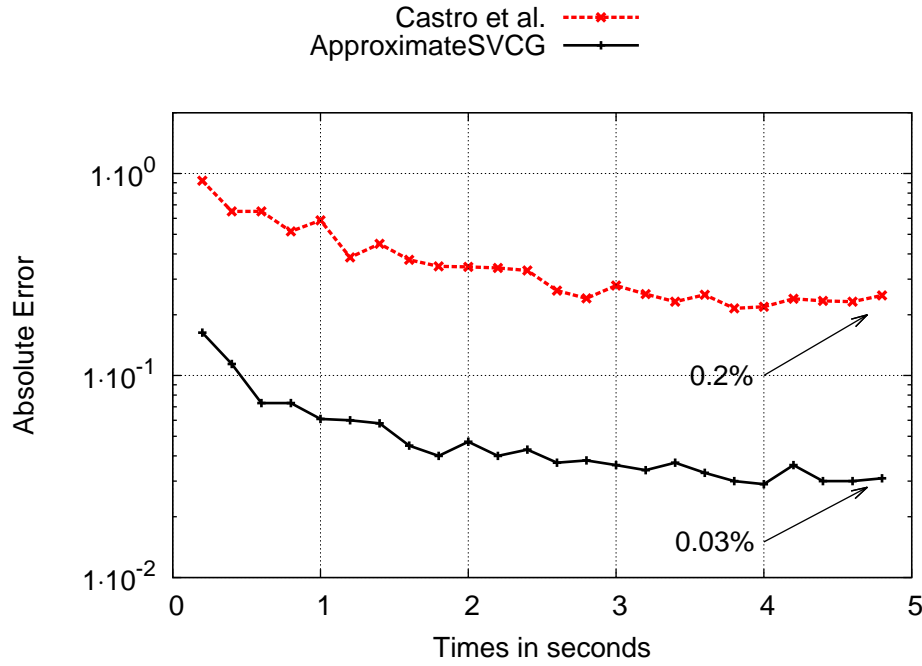


Figure 7.5: Error performance of ApproximateSVCG.

**New insight on Krebs’ 9/11 WTC network:** In the already classic work, Krebs [2002] constructed the 9/11 network from publicly available sources and computed standard centrality metrics to determine the key players in this network. Lindelauf *et al.* instead used their new centrality metric and, similarly to the case of the Jemaah Islamiyah’s network, showed that the Shapley value-based approach delivers qualitatively better insights. However, Lindelauf *et al.*’s analysis focuses only on the network of 19 hijackers with 32 relationships, whereas Krebs reported also a bigger network of 36 nodes and 125 edges (mentioned above) that included accomplices. With FasterSVCG, we were able to analyse this bigger network as well.

Importantly, the five top terrorists (out of 36) identified by FasterSVCG differ from the top five terrorists reported by Lindelauf *et al.* when only the network of 19 terrorist was taken into account—see Table 7.1. This difference is especially striking with respect to **M. Atta**, who was

widely believed to be one of the ring-leaders of the conspiracy [Krebs, 2002] and was positioned the third place in our computations, but was classified in Lindelauf *et al.*'s work only on the 6th place. This clearly shows the importance of developing tools to analyse the biggest possible networks.

## **7.6. Conclusions**

In this chapter, the new algorithm was proposed to compute the Shapley value for the connectivity games of Amer and Gimenez defined over arbitrary graphs. These graphs are used by security services to represent terrorist networks. It was shown that the problem is #P-Hard. Nevertheless, using the new exact algorithm we were able to analyse in moderate time the terrorist network responsible for the 9/11 WTC terrorist attacks. Additionally, in this chapter a new approximate algorithm was proposed that allows for an efficient study of bigger networks.

## 8. The Marginal Contribution Networks for Generalized Coalitional Games

*Generalized characteristic function games* are a variation of characteristic function games, in which the value of a coalition depends not only on the identities of its members, but also on the order in which the coalition is formed. This class of games is a useful abstraction for a number of realistic settings and economic situations, such as modeling relationships in social networks. To date, two main extensions of the Shapley value have been proposed for generalized characteristic function games: the Nowak-Radzik [1994] value and the Sánchez-Bergantiños [1999] value. In this context, the present chapter studies generalized characteristic function games from the point of view of representation and computation. Specifically, we make two key contributions. First, we propose the Generalized Marginal-Contribution Nets that is a fully expressive mechanism that can in natural way represent generalized characteristic function games. Second, in order to facilitate an efficient computation, we propose the polynomial algorithms for computing Nowak-Radzik value and Sánchez-Bergantiños value. Our representation extends the results of Jeong and Shoham [2005] and Elkind et al. [2009a] for characteristic function games and retains their attractive computational properties. Our additional contribution is to use such a representation in order to compute game-theoretic centrality measures defined using either the Nowak-Radzik value or the Sánchez-Bergantiños values that extends path betweenness centrality introduced by Puzis et al. [2007b].

All aforementioned contributions were made exclusively by the author of this dissertation.

### 8.1. Preliminaries

In this chapter we introduce the basic notation for *generalized coalitional games* and also we present two extensions of the Shapley value (Definition 5) to these games.

#### 8.1.1. The Generalized Coalitional Games

For each coalition  $C \in 2^N \setminus \{\emptyset\}$ , denote by  $\Pi(C)$  the set of all possible permutations of the players in  $C$ . Any such permutation will be called an *ordered coalition*. An arbitrary ordered coalition will often be denoted as  $T$ , while the set of all such coalitions will be denoted  $\mathcal{T}$ . That

is,  $\mathcal{T} = \bigcup_{C \in 2^N} \Pi(C)$ . A *generalized characteristic function*  $\nu^g$  is a mapping  $\nu^g : \mathcal{T} \rightarrow \mathbb{R}$ , where it is assumed that  $\nu^g(\emptyset) = 0$ .

**Definition 43 (Generalized coalitional game).** A *generalized coalitional game* consists of a set of players (or agents)  $A = \{a_1, a_2, \dots, a_{|A|}\}$ , and a generalized characteristic function  $\nu^g : \mathcal{T} \rightarrow \mathbb{R}$ , which assigns to each ordered coalition of players  $T \in \mathcal{T}$  a real value (or payoff) indicating its performance, where  $\nu^g(\emptyset) = 0$ . Thus, the generalized coalitional game  $g^g$  in characteristic function form is a pair  $g^g = (A, \nu^g)$ .

A game in generalized characteristic function form is a tuple  $(A, \nu^g)$ , but will sometimes be denoted by  $\nu^g$  alone. Sometimes, we will write  $a_i$  instead of  $(a_i)$  for brevity.

In some cases we will refer to the members of an ordered coalition  $T$  using their names, e.g., write  $T = (a_5, a_2, a_3)$ , while other times we may refer to them using a lower case of the same letter:  $T = (t_1, \dots, t_{|T|})$ , meaning that  $t_i$  is the  $i^{\text{th}}$  agent in  $T$ . Furthermore, given two disjoint ordered coalitions,  $T = (t_1, \dots, t_{|T|}) \in \mathcal{T}$  and  $S = (s_1, \dots, s_{|S|}) \in \mathcal{T}$ , we write  $(T, S)^k$  to denote the ordered coalition that results from inserting  $S$  at the  $k^{\text{th}}$  position in  $T$ . That is,  $(T, S)^k = (t_1, \dots, t_{k-1}, s_1, \dots, s_{|S|}, t_k, \dots, t_{|T|})$ . With a slight abuse of notation, we write  $(T, a_i)^k$  to denote  $(T, (a_i))^k$ . Furthermore, we write  $(a_i, T)$ , and  $(T, a_i)$ , to denote the ordered coalition that results from inserting  $a_i$  to  $T$  as the first player, and the last player, respectively.

Next, we extend the notion of a subset to ordered sets.

**Definition 44 (The subset of ordered sets).** For any two ordered coalitions  $S = (s_1, \dots, s_{|S|}) \in \mathcal{T}$  and  $T = (t_1, \dots, t_{|T|}) \in \mathcal{T}$ , we say that  $T$  is a **subset** of  $S$ , and write  $T \subseteq S$ , if and only if  $T$  is a subsequence of  $S$ , i.e., the following two conditions hold:

— Every member of  $T$  is a member of  $S$ . More formally:

$$\forall t_i \in T, \exists s_k \in S : s_k = t_i.$$

— For any two players,  $t_i, t_j \in T$ , if  $t_i$  appears before  $t_j$  in  $T$ , then  $t_i$  also appears before  $t_j$  in  $S$ . More formally:

$$\forall t_i, t_j \in T : i < j, \exists s_k, s_w \in S : k < w \text{ and } s_k = t_i \text{ and } s_w = t_j.$$

Following convention, we say that  $T$  is a **strict subset** of  $S$ , and write  $T \subset S$  (instead of  $T \subseteq S$ ), if the above two conditions are met, and  $T \neq S$ .

### 8.1.2. The Nowak and Radzik Value

Whereas four axioms defined in Section 2.2 uniquely determine the Shapley value for characteristic function games, the situation is more complex for generalized games, because

a player's marginal contribution (and consequently the *null-player* axiom) depends on where the new player in the coalition is placed. In this respect, Nowak and Radzik [1994] developed an extension of the Shapley value by making perhaps the most natural assumption that the marginal contribution of a player is computed when this player is placed *last* in the coalition. Let us denote this marginal contribution of  $a_i$  to  $T \in \mathcal{T}(N \setminus \{a_i\})$  in game  $\nu^g$  (according to Nowak and Radzik's definition) as  $\text{MC}^{NR}(T, a_i)$ . Then:

$$\text{MC}^{NR}(T, a_i) = \nu^g((T, a_i)) - \nu^g(T). \quad (8.1)$$

In what follows, for any ordered coalition,  $T$ , let  $T(a_i)$  denote the sequence of players in  $T$  that appear before  $a_i$  (if  $a_i \notin T$  then  $T(a_i) = T$ ). For example, given  $T = (a_1, a_3, a_4, a_6)$ , we have  $T(a_4) = (a_1, a_3)$ . Using this notation, we have the following definition.

**Definition 45 (The Nowak-Radzik value).** *In a generalized coalitional game  $(A, \nu^g)$  the Nowak-Radzik value (or NR value for short) for a player  $a_i$  is given by:*

$$\phi_i^{NR}(A, \nu^g) = \frac{1}{|A|!} \sum_{T \in \Pi(A)} \text{MC}^{NR}(T(a_i), a_i) = \mathbb{E}[\text{MC}^{NR}(T(a_i), a_i)]. \quad (8.2)$$

This can be written differently as follows:

$$\phi_i^{NR}(A, \nu^g) = \sum_{C \subseteq A \setminus \{a_i\}} \sum_{T \in \Pi(C)} \frac{(|A| - |T| - 1)!}{|A|!} [\nu^g((T, a_i)) - \nu^g(T)]. \quad (8.3)$$

The NR value is the unique value that satisfies the following “fairness” axioms:

**Efficiency:**  $\sum_{a_i \in A} \phi_i^{NR}(\nu^g) = \frac{1}{|A|!} \sum_{T \in \Pi(A)} \nu^g(T).$

**Null-Player:**  $\forall a_i \in A$ , if  $\nu^g(T) = \nu^g((T, a_i)) \forall T \in \mathcal{T} : a_i \notin T$ , then  $\phi_i^{NR}(\nu^g) = 0$ .

**Additivity:**  $\phi^{NR}(\nu^g + \nu'^g) = \phi^{NR}(\nu^g) + \phi^{NR}(\nu'^g)$  for any two functions,  $\nu^g$  and  $\nu'^g$ .

### 8.1.3. The Sánchez and Bergantiños Value

Sánchez and Bergantiños [1999] developed an alternative extension of the Shapley value based on the definition of the marginal contribution, where, instead of assuming that this player will be placed last, the authors take the average over all possible positions in which the player can be placed:

$$\text{MC}^{SB}(T, a_i) = \frac{1}{(|T| + 1)} \sum_{l=1}^{|T|+1} [\nu^g((T, a_i)^l) - \nu^g(T)]. \quad (8.4)$$

Using above, we can introduce the following definition.



**Definition 46 (The Sánchez-Bergantiños value).** In a generalized coalitional game  $(A, \nu^g)$  the Sánchez-Bergantiños value (or SB value for short) for a player  $a_i$  is given by:

$$\phi_i^{SB}(A, \nu^g) = \frac{1}{|A|!} \sum_{T \in \Pi(A)} \mathbf{MC}^{SB}(T(a_i), a_i) = \mathbb{E}[\mathbf{MC}^{SB}(T(a_i), a_i)]. \quad (8.5)$$

This also can be rewritten differently as follows:

$$\phi_i^{SB}(A, \nu^g) = \sum_{C \subseteq A \setminus \{a_i\}} \sum_{T \in \Pi(C)} \frac{(|A| - |T| - 1)!}{|A|!(|T| + 1)} \sum_{l=1}^{|T|+1} [\nu^g((T, a_i)^l) - \nu^g(T)]. \quad (8.6)$$

The SB value is the unique value that satisfies NR's efficiency and additivity axioms and the following axioms:

**Null-Player** If  $\forall T \in \mathcal{T} \forall l \in \{1, \dots, |T| + 1\} : \nu^g((T, a_i)^l) = \nu^g(T)$ , then  $\phi_i^{SB}(\nu^g) = 0$ .

**Symmetry** If  $\forall T \in \mathcal{T}_{-\{i,j\}} \forall l \in \{1, \dots, |T| + 1\} : \nu^g((T, a_i)^l) = \nu^g((T, a_j)^l)$ , then  $\phi_i^{SB}(\nu^g) = \phi_j^{SB}(\nu^g)$ .

The difference between the NR and SB values is illustrated in the following example:

**Example 7.** Consider a game with an ordered coalition  $T^* \in \Pi(A)$  such that  $\nu^g(T) = 1$  if  $T = T^*$  and  $\nu^g(T) = 0$  otherwise. Then, the average value of the grand coalition, taken over all possible orders, which is  $\frac{1}{|A|!}$ , needs to be distributed among the players. Using the NR value, we get  $\phi_t^{NR}(A) = \frac{1}{|A|!}$ , where  $a_t$  is the last player in the ordered coalition  $T^*$ , and we get  $\phi_i^{NR}(A) = 0$  for all  $a_i \in A \setminus \{a_t\}$ . In contrast, using the SB value, we get  $\phi_i^{SB}(A) = \frac{1}{|A|! \cdot |T^*|} = \frac{1}{|A|! \cdot |A|}$  for all  $a_i \in A$ . As can be seen, in this example, the NR value rewards the last player in the order, whereas the SB value rewards all players equally.

Having introduced the extensions of Shapley value to generalized characteristic function games, in the following section we consider the issue of representation.

## 8.2. Representation

In this section we consider the issue of efficient computation of the NR and SB values. Recall that a straightforward computation of the NR value, using either formulas (8.2) or (8.3), requires iterating over *all ordered coalition*. The same holds for the SB value and formulas (8.5) or (8.6). Now since, for any game of  $n$  players, there are  $\sum_{i=1}^n \binom{n}{i} \times i!$  ordered coalitions, such an approach quickly becomes prohibitive with increasing  $n$ .<sup>1</sup> To tackle this computational problem we introduce in this section a *compact representation scheme* for games with ordered

<sup>1</sup> For instance, the size of the generalized characteristic function for  $n = 5$  is 325, whereas for  $n = 10$  is 9864100.

coalitions and evaluate its properties. We take as our starting point the *marginal contribution nets* (MC-Nets) representation by Ieong and Shoham [2005] (see Section 2.4) and its extension, *read-once marginal contribution nets* (or *read-once MC-Nets*) by Elkind et al. [2009a] (see Section 2.5). This representation has a number of desirable properties: it is fully expressive, concise for many characteristic function games of interest, and facilitates a very efficient technique for computing the Shapley value. Importantly, our representation, which we call *generalized read-once MC-Nets*, also has these desirable properties.

This section is divided into four subsections. In Section 8.2.1, we introduce the *generalized read-once MC-Nets* representation to represent generalized characteristic function games compactly. After that, in Sections 8.2.2 and 8.2.3, we prove that this representation facilitates an efficient computation of the NR value and the SB value, respectively. Finally, in Section 8.3, we present a sample application of generalized read-once MC-Nets to compute game-theoretic centrality based on the NR and SB values.

### 8.2.1. Generalized Read-Once MC-Nets

Having presented the MC-Net representation for characteristic function games in Section 2.4 and read-once MC-Net on Section 2.5, we now present our extension of MC-Nets to generalized characteristic function games. More specifically, we extend the rules so that the formula  $\mathcal{F}$  on the left hand side of a rule is applicable to an *ordered* coalition. We introduce two *atomic formulas*:

$$\begin{aligned} \text{basic atomic formula (BAF)} & \quad \{a_m, \dots, a_n\}, \\ \text{ordered atomic formula (OAF)} & \quad \langle a_i, \dots, a_j \rangle. \end{aligned}$$

The ordered coalition  $T$  meets a basic atomic formula, denoted  $\mathcal{F}_{BAF}$ , if all literals occurring in  $\mathcal{F}_{BAF}$  belong to  $T$ . More formally

$$T \models \mathcal{F}_{BAF} \iff \forall a_i \in \mathcal{F}_{BAF}, a_i \in T.$$

**Example 8.** Let  $N = \{a_1, \dots, a_{10}\}$  and consider  $T = (a_5, a_1, a_4, a_3, a_2)$ . We have:

$$\begin{aligned} T & \models \{a_2, a_5\}, \\ T & \not\models \{a_5, a_2, a_4, a_6\}. \end{aligned}$$

Now, we introduce an ordered atomic formula, denoted  $\mathcal{F}_{OAF}$ . This formula is our crucial extension to MC-Nets; it allows us to fully express the values of all ordered coalitions. Specifically, the ordered coalition  $T$  meets  $\mathcal{F}_{OAF}$  if all literals occurring in  $\mathcal{F}_{OAF}$  appear in  $T$  in the same order as in the  $\mathcal{F}_{OAF}$  formula.

To state this more formally, let us first introduce the notion of an ordered coalition corresponding to  $\mathcal{F}_{OAF}$ . To this end, for every  $\mathcal{F}_{OAF} = \langle a_i, \dots, a_j \rangle$ , the *corresponding ordered coalition* of  $\mathcal{F}_{OAF}$  is simply  $S^{OAF} = (a_i, \dots, a_j)$ . The main difference between the two is that  $a_i$  is a literal in  $\mathcal{F}_{OAF}$ , while it is an agent in  $S^{OAF}$ . Now, given a particular  $\mathcal{F}_{OAF}$  and an ordered coalition  $S^{OAF}$  corresponding to it, the following holds for every  $T \in \mathcal{T}$ :

$$T \models \mathcal{F}_{OAF} \iff S^{OAF} \vec{\subseteq} T.$$

In other words,  $S^{OAF}$  is a subsequence of  $T$ . This has a natural interpretation: when a group of players joins any coalition in a certain order they contribute to it a certain value. Consider the following example:

**Example 9.** Let  $N = \{a_1, \dots, a_{10}\}$  and consider  $T = (a_5, a_1, a_4, a_3, a_2)$ . We have:

$$\begin{aligned} T &\models \langle a_5, a_3, a_2 \rangle, \\ T &\not\models \langle a_5, a_2, a_3 \rangle, \\ T &\not\models \langle a_5, a_3, a_2, a_6 \rangle. \end{aligned}$$

Now, we are able to formalize our representation which we call the *generalized read-once MC-Nets*. Specifically, in this representation, a rule, called *generalized read-once rule*, is as follows:

$$\mathcal{F} \rightarrow V, \tag{8.7}$$

where  $V$  is a real value and the syntax of the *generalized formula*  $\mathcal{F}$  is:

$$\begin{aligned} A &:= \mathcal{F}_{BAF} \mid \mathcal{F}_{OAF}, \\ \mathcal{F} &:= (\mathcal{F}) \mid \neg \mathcal{F} \mid \mathcal{F} \wedge \mathcal{F} \mid \mathcal{F} \vee \mathcal{F} \mid \mathcal{F} \oplus \mathcal{F} \mid A, \end{aligned}$$

and each literal can only appear once in the formula. This generalized formula is a binary rooted tree with atomic formulas on leafs and internal nodes labeled with one of the following logical connectives: conjunction ( $\wedge$ ), disjunction ( $\vee$ ) or exclusive disjunction ( $\oplus$ ). The logical connectives have the following standard interpretation:

$$\begin{aligned} T \models \neg \mathcal{F} &\iff T \not\models \mathcal{F}, \\ T \models \mathcal{F}_1 \wedge \mathcal{F}_2 &\iff T \models \mathcal{F}_1 \text{ and } T \models \mathcal{F}_2, \\ T \models \mathcal{F}_1 \vee \mathcal{F}_2 &\iff T \models \mathcal{F}_1 \text{ or } T \models \mathcal{F}_2, \\ T \models \mathcal{F}_1 \oplus \mathcal{F}_2 &\iff T \models (\mathcal{F}_1 \vee \mathcal{F}_2) \wedge \neg(\mathcal{F}_1 \wedge \mathcal{F}_2). \end{aligned}$$

**Example 10.** Let  $N = \{a_1, \dots, a_{10}\}$  and consider  $T = (a_5, a_1, a_4, a_3, a_2)$ . We have:

$$\begin{aligned} T &\models (\{a_2, a_5\} \wedge \langle a_1, a_3 \rangle) \vee \{a_4, a_7, a_9\}, \\ T &\not\models \langle a_4, a_5, a_3 \rangle \oplus (\{a_1\} \wedge \neg\{a_2\}). \end{aligned}$$

**Definition 47 (Generalized Read-Once MC-Nets).** The generalized coalitional game  $(N, \nu^g)$  in the form of Generalized Read-once Marginal Contribution Network is a pair  $(A, \mathcal{GR})$ , where  $A$  is a set of players, and  $\mathcal{GR}$  is a finite set of generalized rules. The value  $\nu^g(T)$  of an ordered coalition  $T$  is defined as the sum of all  $V$  from the generalized read-once rules that are met by  $T$ . More formally:

$$\nu^g(T) = \sum_{\mathcal{GR} \ni \mathcal{F} \rightarrow V: T \models \mathcal{F}} V. \quad (8.8)$$

The following example presents a simple application of generalized read-once MC-Nets to the scheduling problem:

**Example 11.** Let  $G = (V, E)$  be a directed acyclic graph representing a task-based domain, where each vertex in  $V = \{v_1, \dots, v_n\}$  represents a task. Each edge  $(v_i, v_j) \in E$  illustrates that task  $v_i$  has to be completed before task  $v_j$ . We say that the job is done if all tasks of a given graph  $G$  are completed, in which case we get some revenue. Next we show how each such graph can be represented by a single generalized read-once rule, where agents represent tasks. In particular, each formula in the rule represents an edge in  $G$ , as follows:

$$\bigwedge_{(a_i, a_j) \in E} \langle a_i, a_j \rangle \rightarrow \text{revenue of the job defined by } G.$$

An ordered coalition  $T$  represents the workforce that can finish tasks in a given order. The total revenue of  $T$  is then equal to the sum of revenues of all jobs feasible by  $T$ .

Having defined our representation formally, we now evaluate its properties. We start with expressiveness:

**Proposition 5 (Expressiveness).** Every ordered coalitional game can be expressed using generalized read-once MC-Nets.

*Proof.* For any arbitrary generalized characteristic function  $\nu^g$ , we show how to construct a set of rules  $\mathcal{GR}$  such that Equation (8.8) holds for every  $T \in \mathcal{T}$ .

We will use only OAF to define a recursive set starting from singleton sequences and empty set  $\mathcal{GR}_0 = \emptyset$ . Let  $S^{OAF}$  be the ordered coalition corresponding to  $\mathcal{F}_{OAF}$ , and let  $|\mathcal{F}_{OAF}|$  be the size of this formula.

$$\begin{aligned} \mathcal{GR}_1 &= \{ \langle a_i \rangle \rightarrow \nu^g((a_i)) \} & \forall a_i \in N, \\ \mathcal{GR}_n &= \{ \mathcal{F}_{OAF} \rightarrow \nu^g(S^{OAF}) - \sum_{k=1}^{n-1} \sum_{\mathcal{GR}_k \ni \mathcal{F} \rightarrow V: S^{OAF} \models \mathcal{F}} V \} & \forall |\mathcal{F}_{OAF}| = n. \end{aligned}$$

Finally, we get  $\mathcal{GR} = \bigcup_i \mathcal{GR}_i$ . ■

In the above proof of Proposition 5, it emerges that only OAFs are necessary in order to guarantee that the generalized MC-Nets are fully expressive. In the following basic example we present that in some games the BAF gives us desirable succinctness.

**Example 12.** *Let  $N = \{a_1, \dots, a_n\}$  be the set of tasks. In order to get profit  $V$ , all tasks must be done in arbitrary chosen order. It is clear that representing this problem using only OAF requires  $n!$  number of formulas:*

$$\begin{aligned} \langle a_1, \dots, a_n \rangle &\rightarrow V, \\ &\vdots \\ \langle a_n, \dots, a_1 \rangle &\rightarrow V. \end{aligned}$$

*By introducing BAF we can get exponentially more concise notation:  $\{a_1, \dots, a_n\} \rightarrow V$ .*

It is worth making some observations about the conciseness of our representation scheme. Our rule-based representation scheme is ultimately based on propositional formulae, and the use of these formulae is ultimately to define a Boolean function. The use of propositional formulae to define Boolean functions is very standard: a very fundamental result in the theory of propositional logic is that any Boolean function of  $n$  Boolean arguments can be represented by a propositional formula in which the variables correspond to these arguments. For many Boolean functions, it might be possible to obtain *compact* propositional formulae to represent them, but crucially, this will not always be possible; some Boolean functions will require formulae of size exponential in the number of variables (see, e.g., Boppana and Sipser [1990]). It is in fact trivial to produce examples of characteristic functions whose representation in our framework is exponentially more succinct than explicitly listing the value of every coalition, but from the above discussion, it also follows that our representation cannot *guarantee* compactness in all cases. Finally, note that in the worst case, we can represent a characteristic function with one rule for every possible input, which essentially corresponds to the idea of listing the value of every coalition. Thus, in the worst case, our representation scheme is no worse than simply listing the value of every input. More formally:

**Corollary 1 (Conciseness).** *Compared to the generalized function game representation, generalized read-once MC-nets are at least as concise, and for certain games exponentially more compact.*

### 8.2.2. Computing the Nowak-Radzik Value with Generalized Read-Once MC-Nets

In this subsection, we discuss the computational properties of the generalized read-once MC-Net representation with respect to calculating the Nowak-Radzik value (i.e., the NR value).

We start our computational analysis with some observations about atomic formulas. First, observe that all players occurring in some  $\mathcal{F}_{BAF}$  are indistinguishable. Furthermore, taking into account that each player can appear in at most one formula,  $\phi^{NR}$  will assign to these players the same values. The case of  $\mathcal{F}_{OAF}$  is different. Here, the NR value for all players except the last one is zero (see Example 7).

---

**Algorithm 22: (COMP\_NR) Computing the Nowak-Radzick value**

---

**Input:**  $N = \{a_1, \dots, a_n\}, (\mathcal{F}_1 \rightarrow V_1), \dots, (\mathcal{F}_r \rightarrow V_r)$   
**Output:**  $\phi_i^{NR}$  for all  $a_i \in N$

```

1 for  $a_i \in N$  do  $\phi_i^{NR} \leftarrow 0$ ; for  $j \leftarrow 1$  to  $r$  do
2    $m \leftarrow |\mathcal{F}_j|$ ;
3    $(A, B, T, F) \leftarrow Sh^{NR}$ ; // the Pseudo code of  $Sh^{NR}$  is in Appendix B
4   foreach  $a_i \in X_{\mathcal{F}_j}$  do
5     if  $\exists_k A_{k,i} \neq 0$  then
6        $v_i \leftarrow \frac{1}{m!} \sum_{k=0}^{m-1} (m-k-1)! A_{k,i}$ ;
7     else if  $\exists_k B_{k,i} \neq 0$  then
8        $v_i \leftarrow -\frac{1}{m!} \sum_{k=0}^{m-1} (m-k-1)! B_{k,i}$ ;
9     else
10       $v_i \leftarrow 0$ ;
11     $\phi_i^{NR} \leftarrow \phi_i^{NR} + V_j \cdot v_i$ ;

```

---

**Theorem 6.** *Given a generalized coalitional game  $(N, \mathcal{GR})$ , Algorithm 22 computes the Nowak-Radzick value  $(\phi_i^{NR})$  for all  $a_i \in N$  in polynomial time.*

*Proof.* We start by introducing necessary notation, some of which is in the pseudo code of Algorithm 22. For each formula  $\mathcal{F}$ , we denote by  $X_{\mathcal{F}}$  the sets of players occurring in  $\mathcal{F}$ , and by  $|\mathcal{F}|$  the size of  $X_{\mathcal{F}}$ , and by  $\mathcal{T}_{\mathcal{F}}$  the set of all ordered coalitions within the set  $X_{\mathcal{F}}$ , i.e.,  $\mathcal{T}_{\mathcal{F}} = \bigcup_{C \subseteq X_{\mathcal{F}}} \Pi(C)$ . For all  $a_i \in X_{\mathcal{F}}$  and  $k \in \{0, \dots, n\}$ , we define the following quantities:

$$\begin{aligned}
A_{k,i}(\mathcal{F}) &= |\{T \in \mathcal{T}_{\mathcal{F}} : |T| = k, a_i \notin T, T \not\models \mathcal{F}, (T, a_i) \models \mathcal{F}\}|, \\
B_{k,i}(\mathcal{F}) &= |\{T \in \mathcal{T}_{\mathcal{F}} : |T| = k, a_i \notin T, T \models \mathcal{F}, (T, a_i) \not\models \mathcal{F}\}|, \\
T_k(\mathcal{F}) &= |\{T \in \mathcal{T}_{\mathcal{F}} : |T| = k, T \models \mathcal{F}\}|, \\
F_k(\mathcal{F}) &= |\{T \in \mathcal{T}_{\mathcal{F}} : |T| = k, T \not\models \mathcal{F}\}|.
\end{aligned}$$

The first quantity  $A_{k,i}(\mathcal{F})$  is the number of ordered coalitions from  $\mathcal{T}_{\mathcal{F}}$  of a given size  $k$  (i.e., containing  $k$  players) to which the addition of player  $a_i$  at the end of the coalition causes the formula  $\mathcal{F}$  to become satisfied. In contrast, the second quantity  $B_{k,i}$  is the number of coalitions to which adding player  $a_i$  causes the formula  $\mathcal{F}$  to become unsatisfied. The last two quantities,  $T_k(\mathcal{F})$  and  $F_k(\mathcal{F})$ , count the number of ordered coalition being satisfied, and not satisfied, respectively, by the formula  $\mathcal{F}$ . These quantities can be computed recursively as follows. Let

us first focus on the case where  $\mathcal{F}$  is an atomic formula. This case can be divided further into two cases:

- $\mathcal{F} = \{a_1, \dots, a_r\}$ : Here,  $A_{k,i} = (r-1)!$  if  $k = r-1$  and  $i \in X_{\mathcal{F}}$ , otherwise  $A_{k,i} = 0$ .  $B_{k,i}$  always equals 0.  $T_k = r!$  if  $k = r$ , otherwise  $T_k = 0$ . Finally,  $F_k = k!$  if  $k < r$ , otherwise  $F_k = 0$ .
- $\mathcal{F} = \langle a_1, \dots, a_r \rangle$ : Here  $A_{k,i} = 1$  if  $k = r-1$  and  $i = r$ , otherwise  $A_{k,i} = 0$ .  $B_{k,i}$  always equals 0.  $T_k = 1$  if  $k = r$ , otherwise  $T_k = 0$ . Finally,  $F_k = k! - T_k$  if  $k \leq r$ , otherwise  $F_k = 0$ .

Having discussed the case where  $\mathcal{F}$  is an atomic formula, we now consider the case where  $\mathcal{F}$  contains logical connectives. If such a connective involves two subformulas  $\mathcal{F}_1$  and  $\mathcal{F}_2$ , then we know that  $a_i$  can appear in  $\mathcal{F}_1$  or  $\mathcal{F}_2$ , but not in both. We also know that logical connectives used in the generalized read-once MC-Nets are symmetric, so we will consider only one case,  $a_i \in \mathcal{F}_1$ ; the other is analogous.

In our analysis, for every two disjoint ordered coalitions,  $T_1$  and  $T_2$ , we define a *conflation* of  $T_1$  and  $T_2$  as an ordered coalition  $T$  such that (1) every agent in  $T$  appears in  $T_1$  or  $T_2$ , and (2) the order of the agents in  $T_1$  is retained in  $T$ , and the order of those in  $T_2$  is also retained in  $T$ . The set of every conflation of  $T_1$  and  $T_2$  will be denoted  $T_1 \times T_2$ . More formally:

$$T \in T_1 \times T_2 \iff X_T = X_{T_1} \cup X_{T_2} \wedge T' \preceq T \wedge T_2 \preceq T,$$

where  $X_T$  denotes a set of players in  $T$ . We note that if  $T_1 \models \mathcal{F}$  and  $X_{\mathcal{F}} \cap T_2 = \emptyset$  then  $T \models \mathcal{F} \forall T \in (T_1 \times T_2)$ .

**Proposition 6.** *Two ordered coalitions  $T_1$  and  $T_2$  can be conflated in  $\binom{|T_1|+|T_2|}{|T_2|}$  ways.*

*Proof.* Consider an sequence of  $|T_1| + |T_2|$  empty slots, where each slot can take an agent. Out of all these slots, choose any  $|T_2|$  slots, and place in them the agents from  $T_2$  while retaining their order. As for the remaining slots, place in them the agents from  $T_1$  while retaining their order. Clearly, every choice of the  $|T_2|$  slots results in a unique conflation, and every conflation can be constructed by exactly one such choice of the  $|T_2|$  slots. The number of conflations is then  $\binom{|T_1|+1+|T_2|-1}{|T_2|}$ . ■

Next, we show how to compute  $A_{k,i}, B_{k,i}, T_k$  and  $F_k$ . We will do this for each of the possible logical connective (i.e.,  $\neg, \wedge, \vee$  and  $\oplus$ ) separately:

- $\mathcal{F} = \neg \mathcal{F}_1$ : In this case, the negation swaps quantities as follows. For all  $i$  and  $k$  we have  $A_{k,i}(\mathcal{F}) = B_{k,i}(\mathcal{F}_1)$ ,  $B_{k,i}(\mathcal{F}) = A_{k,i}(\mathcal{F}_1)$ ,  $T_k(\mathcal{F}) = F_k(\mathcal{F}_1)$ , and  $F_k(\mathcal{F}) = T_k(\mathcal{F}_1)$ .
- $\mathcal{F} = \mathcal{F}_1 \wedge \mathcal{F}_2$ . In this case, let  $T \in T_1 \times T_2$ , where  $T_1 \in \Pi(X_{\mathcal{F}_1}), |T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2}), |T_2| = k - s$ . We have  $(T, a_i) \models \mathcal{F}$  if and only if  $(T_1, a_i) \models \mathcal{F}_1$  and

$T_2 \models \mathcal{F}_2$ . In this case,  $T \not\models \mathcal{F}$  if and only if  $T_1 \not\models \mathcal{F}_1$ . Consequently, using Proposition 6 we get:

$$A_{k,i} = \sum_{s=0}^k \binom{k}{k-s} A_{s,i}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2).$$

Furthermore, we have  $T \models \mathcal{F}$  if and only if  $T_1 \models \mathcal{F}_1$  and  $T_2 \models \mathcal{F}_2$ . Additionally in this case we have:  $(T, a_i) \not\models \mathcal{F} \Leftrightarrow (T_1, a_i) \not\models \mathcal{F}_1$ . Consequently,

$$B_{k,i} = \sum_{s=0}^k \binom{k}{k-s} B_{s,i}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2).$$

Eventually, for  $T_k$  and  $F_k$ , we have:

$$\begin{aligned} T_k &= \sum_{s=0}^k \binom{k}{k-s} T_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2), \\ F_k &= \sum_{s=0}^k \binom{k}{k-s} (F_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2) + F_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + T_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2)). \end{aligned}$$

—  $\mathcal{F} = \mathcal{F}_1 \vee \mathcal{F}_2$ . In this case, let  $T \in T_1 \times T_2$ , where  $T_1 \in \Pi(X_{\mathcal{F}_1})$ ,  $|T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2})$ ,  $|T_2| = k - s$ . We know that  $T \not\models \mathcal{F}$  if and only if  $T_1 \not\models \mathcal{F}_1$  and  $T_2 \not\models \mathcal{F}_2$ . Furthermore, in this case it holds that  $(T, a_i) \models \mathcal{F} \Leftrightarrow (T_1, a_i) \models \mathcal{F}_1$ . Thus,

$$A_{k,i} = \sum_{s=0}^k \binom{k}{k-s} A_{s,i}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2).$$

Analogously, we have  $(T, a_i) \not\models \mathcal{F}$  if and only if  $(T_1, a_i) \not\models \mathcal{F}_1$  and  $T_2 \not\models \mathcal{F}_2$ , and in this case it holds that:  $T \models \mathcal{F} \Leftrightarrow T_1 \models \mathcal{F}_1$ . Consequently,

$$B_{k,i} = \sum_{s=0}^k \binom{k}{k-s} B_{s,i}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2).$$

Finally, for  $T_k$  and  $F_k$ , we have:

$$\begin{aligned} T_k &= \sum_{s=0}^k \binom{k}{k-s} (T_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + F_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + T_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2)), \\ F_k &= \sum_{s=0}^k \binom{k}{k-s} F_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2). \end{aligned}$$

—  $\mathcal{F} = \mathcal{F}_1 \oplus \mathcal{F}_2$ : Let  $T = T_1 \times T_2$  such that  $T_1 \in \Pi(X_{\mathcal{F}_1})$ ,  $|T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2})$ ,  $|T_2| = k - s$ . Now, the ordered coalition  $T$  can contribute to  $A_{k,i}$  in two ways. Firstly, when  $T_1 \models \mathcal{F}_1$  and  $T_2 \models \mathcal{F}_2$  and  $(T_1, a_i) \not\models \mathcal{F}_1$ . Secondly, when  $T_1 \not\models \mathcal{F}_1$  and  $T_2 \not\models \mathcal{F}_2$  and  $(T_1, a_i) \models \mathcal{F}_1$ . Thus:

$$\begin{aligned} A_{k,i} &= \sum_{s=0}^k \binom{k}{k-s} (B_{s,i}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + A_{s,i}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2)), \\ B_{k,i} &= \sum_{s=0}^k \binom{k}{k-s} (B_{s,i}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2) + A_{s,i}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2)), \\ T_k &= \sum_{s=0}^k \binom{k}{k-s} (T_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2) + F_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2)), \\ F_k &= \sum_{s=0}^k \binom{k}{k-s} (T_s(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + F_s(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2)). \end{aligned}$$



The above four quantities are computed recursively in the function  $Sh^{NR}$  which is invoked in line 4 of Algorithm 22 (the pseudo code of  $Sh^{NR}$  can be found in Appendix A).

Note that  $T_k$  and  $F_k$  are only used to compute  $A_{k,i}$  and  $B_{k,i}$ , which are then used in lines 6 and 8 of Algorithm 22 to compute  $\phi^{NR}$ . Having proved the correctness of computing  $A_{k,i}$  and  $B_{k,i}$ , it remains to prove that the way they are used in lines 6 and 8 correctly computes  $\phi^{NR}$ . It suffices to prove this for a single rule  $\mathcal{F} \rightarrow 1$ , as the extension to multiple rules is straight forward in the MC-Net representation. To this end, some player  $a_i$  can contribute to any ordered coalition  $T$  counted in  $A_{k,i}$ , or  $B_{k,i}$  (one of these quantities is zero), if it appears in a random permutation right after the players from  $T$ . The remaining players not occurring in  $T$  can be arranged in  $(m - k - 1)!$  ways (we have  $|X_{\mathcal{F}}| = m$ ). In case  $a_i$  contributes to  $A_{k,i}$ , we have

$$\phi_i^{NR} = \frac{1}{m!} \sum_{k=0}^{m-1} (m - k - 1)! A_{k,i}.$$

In the second case  $a_i$  contributes to  $B_{k,i}$ ; we get

$$\phi_i^{NR} = -\frac{1}{m!} \sum_{k=0}^{m-1} (m - k - 1)! B_{k,i}.$$

These are exactly lines 6 and 8 of Algorithm 22. This concludes the proof of correctness of Algorithm 22. In terms of runtime, it is clear from the pseudo codes (in Algorithm 22 and Appendix B) that the total number of operations is polynomial in the number of agents. This concludes the proof of Theorem 6. ■

### 8.2.3. Computing the Sánchez-Bergantiños Value with Generalized Read-Once MC-Nets

In this subsection, we discuss the computational properties of the generalized read-once MC-Net representation with respect to calculating the Sánchez-Bergantiños value (i.e., the SB value). We start our computational analysis with some observations about atomic formulas. First, observe that all players occurring in some  $\mathcal{F}_{BAF}$  are indistinguishable. Furthermore, taking into account that each player can appear in at most one formula,  $\phi^{SB}$  will assign to these players the same values. Similarly, for the case of  $\mathcal{F}_{OAF}$ , the SB value is the same for all players due to the symmetry axiom (see Example 7).

Note that the SB value involves a different notion of marginal contribution. Whereas computing quantities  $T_k$  and  $F_k$  remains the same, we need to redefine  $A_{k,i}$  and  $B_{k,i}$ .

$$\begin{aligned} A_{k,i,l}(\mathcal{F}) &= |T \in \mathcal{T}_{\mathcal{F}} : |T| = k, a_i \notin T, T \not\models \mathcal{F}, (T, a_i)^l \models \mathcal{F}|, \\ B_{k,i,l}(\mathcal{F}) &= |T \in \mathcal{T}_{\mathcal{F}} : |T| = k, a_i \notin T, T \models \mathcal{F}, (T, a_i)^l \not\models \mathcal{F}|. \end{aligned}$$

The first quantity  $A_{k,i,l}(\mathcal{F})$  is the number of ordered coalitions of a given size  $k$  for which adding player  $a_i$  on position  $l$  causes  $\mathcal{F}$  to become satisfied.  $B_{k,i,l}$  is the number of coalitions in which adding the player  $a_i$  on position  $l$  causes  $\mathcal{F}$  to become unsatisfied.

---

**Algorithm 23: (COMP\_SB) Computing the Sánchez-Bergantiños value**

---

**Input:**  $N = \{a_1, \dots, a_n\}, (\mathcal{F}_1 \rightarrow V_1), \dots, (\mathcal{F}_r \rightarrow V_r)$   
**Output:**  $\phi_i^{SB}$  for all  $a_i \in N$

```

1 for  $a_i \in N$  do  $\phi_i^{SB} \leftarrow 0$ ; for  $j \leftarrow 1$  to  $r$  do
2    $m \leftarrow |\mathcal{F}_j|$ ;
3    $(A, B, T, F) \leftarrow Sh^{SB}(\mathcal{F}_j)$ ; // the Pseudo code of  $Sh^{SB}$  is in Appendix B
4   foreach  $a_i \in X_{\mathcal{F}_j}$  do
5     if  $\exists_{k,l} A_{k,i,l} \neq 0$  then
6        $v_i \leftarrow \frac{1}{m!} \sum_{k=0}^{m-1} \frac{1}{k+1} \sum_{l=1}^{k+1} (m-k-1)! A_{k,i,l}$ ;
7     else if  $\exists_{k,l} B_{k,i,l} \neq 0$  then
8        $v_i \leftarrow -\frac{1}{m!} \sum_{k=0}^{m-1} \frac{1}{k+1} \sum_{l=1}^{k+1} (m-k-1)! B_{k,i,l}$ ;
9     else
10       $v_i \leftarrow 0$ ;
11     $\phi_i^{SB} \leftarrow \phi_i^{SB} + V_j \cdot v_i$ ;
```

---

**Theorem 7.** Given a generalized coalitional game  $(N, \mathcal{GR})$ , Algorithm 23 computes the Sánchez-Bergantiños value ( $\phi_i^{SB}$  for all  $a_i \in N$ ) in polynomial time.

*Proof.* As for  $T_k$  and  $F_k$ , the computation is exactly the same as in the proof of Theorem 6. Therefore, we only need to show how to calculate  $A_{k,i,l}$  and  $B_{k,i,l}$  in polynomial time. We will start from atomic formulas:

- $\mathcal{F} = \{a_1, \dots, a_r\}$ : In this case  $A_{k,i,l} = (r-1)!$  if  $k = r-1$ ,  $i \in X_{\mathcal{F}}$  and  $l \in \{1, \dots, r\}$ , otherwise,  $A_{k,i,l} = 0$ . On the other hand,  $B_{k,i,l} = 0$  for all  $k, i$  and  $l$ .
- $\mathcal{F} = \langle a_1, \dots, a_r \rangle$ : Here  $A_{k,i,l} = 1$  if  $k = r-1$ ,  $i \in X_{\mathcal{F}}$  and  $l = i$ , otherwise  $A_{k,i,l} = 0$ . On the other hand,  $B_{k,i,l}$  is always equal to 0.

In the next part of proof we will use following proposition.

**Proposition 7.** The number of conflations of two ordered coalitions  $T_1$  and  $T_2$  such that  $s$  players from  $T_2$  go before the  $l$ -th member of  $T_1$  equals

$$\binom{l+s-1}{s} \binom{(|T_1|-l+1+1)+(|T_2|-s)-1}{|T_2|-s}.$$

*Proof.* Simply use combination with repetition twice. ■

Next, we show how to compute  $A_{k,i}, B_{k,i}$  for each of the possible logical connective (i.e.,  $\neg$ ,  $\wedge$ ,  $\vee$  and  $\oplus$ ) separately. We will assume that  $a_i \in X_{\mathcal{F}_1}$ .

- $\mathcal{F} = \neg \mathcal{F}_1$ : Here, the negation causes a swap between sets. That is, we have  $A_{k,i,l}(\mathcal{F}) = B_{k,i,l}(\mathcal{F}_1)$  and  $B_{k,i,l}(\mathcal{F}) = A_{k,i,l}(\mathcal{F}_1)$  for all  $k, i, l$ .
- $\mathcal{F} = \mathcal{F}_1 \wedge \mathcal{F}_2$ : Let  $T = T_1 \times T_2$ , where  $T_1 \in \Pi(X_{\mathcal{F}_1})$ ,  $|T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2})$ ,  $|T_2| = k - s$ . We have  $(T, a_i)^l \models \mathcal{F}$  if and only if  $(T_1, a_i)^m \models \mathcal{F}_1$  for each  $m \leq l$  when  $l - m$  members of coalition  $T_2$  go before the  $m$ -th member of coalition  $T_1$  during the conflation operation and  $T_2 \models \mathcal{F}_2$ . In this case,  $T \not\models \mathcal{F}$  if and only if  $T_1 \not\models \mathcal{F}_1$ . Consequently, using Proposition 7 we get:

$$A_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} A_{s,i,m}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2).$$

Furthermore, we have  $T \models \mathcal{F}$  if and only if  $T_1 \models \mathcal{F}_1$  and  $T_2 \models \mathcal{F}_2$ . Additionally in this case it holds that:  $(T, a_i)^l \not\models \mathcal{F} \Leftrightarrow (T_1, a_i)^m \not\models \mathcal{F}_1$  for  $m \leq l$ , when  $l - m$  members of coalition  $T_2$  go before the  $m$ -th member of coalition  $T_1$  during the conflation operation and  $T_2 \not\models \mathcal{F}_1$ . Again, using Proposition 7 we get:,

$$B_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} B_{s,i,m}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2).$$

- $\mathcal{F} = \mathcal{F}_1 \vee \mathcal{F}_2$ : Once again let  $T = T_1 \times T_2$ , where  $T_1 \in \Pi(X_{\mathcal{F}_1})$ ,  $|T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2})$ ,  $|T_2| = k - s$ . We know that  $T \not\models \mathcal{F}$  if and only if  $T_1 \not\models \mathcal{F}_1$  and  $T_2 \not\models \mathcal{F}_2$ . Furthermore, in this case it holds that  $(T, a_i)^l \models \mathcal{F} \Leftrightarrow (T_1, a_i)^m \models \mathcal{F}_1$  for  $m \leq l$ . Thus,

$$A_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} A_{s,i,m}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2).$$

Analogously, we have  $(T, a_i)^l \not\models \mathcal{F}$  if and only if  $(T_1, a_i)^m \not\models \mathcal{F}_1$  for  $m \leq l$  when  $l - m$  members of coalition  $T_2$  go before the  $m$ -th member of coalition  $T_1$  during the conflation operation, and  $T_2 \not\models \mathcal{F}_2$ . Furthermore, in this case it holds that:  $T \models \mathcal{F} \Leftrightarrow T_1 \models \mathcal{F}_1$ . Consequently,

$$B_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} B_{s,i,m}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2).$$

- $\mathcal{F} = \mathcal{F}_1 \oplus \mathcal{F}_2$ : Let  $T = T_1 \times T_2$  such that  $T_1 \in \Pi(X_{\mathcal{F}_1})$ ,  $|T_1| = s$  and  $T_2 \in \Pi(X_{\mathcal{F}_2})$ ,  $|T_2| = k - s$ . Now, the ordered coalition  $T$  can contribute to  $A_{k,i,l}$  in two ways. Firstly, when  $T_1 \models \mathcal{F}_1$  and  $T_2 \models \mathcal{F}_2$  and  $(T_1, a_i)^m \not\models \mathcal{F}_1$ . Secondly, when  $T_1 \not\models \mathcal{F}_1$  and  $T_2 \not\models \mathcal{F}_2$  and  $(T_1, a_i)^m \models \mathcal{F}_1$ , so

$$A_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} (B_{s,i,m}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2) + A_{s,i,m}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2)).$$

$$B_{k,i,l} = \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} (B_{s,i,m}(\mathcal{F}_1) F_{k-s}(\mathcal{F}_2) + A_{s,i,m}(\mathcal{F}_1) T_{k-s}(\mathcal{F}_2)).$$

The above two quantities (together with  $\mathbb{T}_k$  and  $\mathbb{F}_k$ ) are computed recursively in the function  $Sh^{SB}$  which is invoked in line 4 of Algorithm 23 (the pseudo code of  $Sh^{SB}$  can be found in Appendix B).

Note that  $\mathbb{T}_k$  and  $\mathbb{F}_k$  are only used to compute  $A_{k,i,l}$  and  $B_{k,i,l}$ , which are then used in lines 6 and 8 of Algorithm 23 to compute  $Sh^{SB}$ . Having proved the correctness of computing  $A_{k,i,l}$  and  $B_{k,i,l}$ , it remains to prove that the way they are used in lines 6 and 8 correctly computes  $Sh^{SB}$ . It suffices to prove this for a single rule  $\mathcal{F} \rightarrow 1$ , as the extension to multiple rules is straight forward in the MC-Net representation. To this end, some player  $a_i$  can contribute to any ordered coalition  $T$  counted in  $A_{k,i,l}$ , or  $B_{k,i,l}$ , if it appears in a permutation right after the  $(l-1)^{th}$  player from  $T$ . The remaining players (i.e., those not belonging to  $T$ ) can be arranged in  $(m-k-1)!$  ways (we have  $|X_{\mathcal{F}}| = m$ ). In case  $a_i$  contributes to  $A_{k,i,l}$ , we have

$$\phi_i^{SB} = \frac{1}{m!} \sum_{k=0}^{m-1} \frac{1}{k+1} \sum_{l=1}^{k+1} (m-k-1)! A_{k,i,l}.$$

In the second case, when  $a_i$  contributes to  $B_{k,i,l}$ , we get

$$\phi_i^{SB} = -\frac{1}{m!} \sum_{k=0}^{m-1} \frac{1}{k+1} \sum_{l=1}^{k+1} (m-k-1)! B_{k,i,l}.$$

These are exactly lines 6 and 8 of Algorithm 23. This concludes the proof of correctness of Algorithm 23. In terms of runtime, it is clear from the pseudo codes (in Algorithm 23 and Appendix B) that the total number of operations is polynomial in the number of agents. This concludes the proof of Theorem 7. ■

Finally, it is worth making a few remarks about the problem of *generating* a representation of a game in our framework. We know that given a representation of a game in our framework, we can for example compute the NR and SB values in polynomial time. But how hard is it to come up with a representation of a game in our framework?

Suppose we are given a black box representation of a characteristic function, i.e., an oracle for the function that can be queried in unit time to determine the value of any ordered coalition. Then, it is not hard to see that it will require time exponential in the number of players to produce a representation of it in our framework. This is because we cannot avoid having to query the oracle for the value of every possible ordered coalition, and there are clearly exponentially many of them.

Another obvious question is whether other representations can easily be “translated” to our representation. Here, we have better news. For example, a representation of a game in Jeong and Shoham’s MC-Nets representation can be transformed into our representation in polynomial time: MC-Net rules correspond directly to rules in our framework requiring only basic atomic formulae (BAFs). In the same way, the subgraph representation introduced by Deng and Papadimitriou [1994] (which is, in fact, a special case of the MC-Net representation)

can also be translated into our framework in polynomial time. Each edge of the graph translates to a rule with a BAF containing only two players.

It is worth noting that we are not aware of any other attempts in the literature to develop representations for generalized coalitional games. If such were developed, then it would be interesting to consider the problem of translating between representations.

### 8.3. Computing Generalized Game-Theoretic Betweenness Centrality

As mentioned in the introduction, it is possible to use the game-theoretic solution concepts, including the Nowak-Radzik and Sánchez-Bergantiños values, to measure centrality of nodes in networks [del Pozo et al., 2011]. In this section, we show how to extend the classic notion of betweenness centrality to generalized game-theoretic betweenness centrality which builds upon generalized characteristic function games and the NR and SB values. After that, we show that it is possible to use generalized read-once MC-Nets to compute the new centrality in time polynomial in the number of shortest paths in the network.

While betweenness centrality (Definition 16) is focused on the role in the network played by individual nodes, and group betweenness centrality (Definition 22) is designed to measure the power of the set of nodes, the measure proposed by Puzis et al. [2007b] captures the notion of the sequences of nodes. Their measure, called *path betweenness centrality*, evaluates sequences of nodes and counts the proportion of shortest paths that traverse along a given sequence of nodes in an appropriate order. Formally:

**Definition 48 (Path betweenness centrality).** *Given a graph  $G = (V, E)$ , we define the set of all ordered coalitions of nodes in the graph, i.e.,  $\mathcal{T} = \bigcup_{C \in 2^V} \Pi(C)$ . The path betweenness centrality of the ordered group of nodes  $T$  is defined as a function  $c_{pb} : \mathcal{T} \rightarrow \mathbb{R}$  such that:*

$$c_{pb}(T) = \sum_{\substack{s \notin T \\ t \notin T}} \frac{\sigma_{st}(T)}{\sigma_{st}}, \quad (8.9)$$

where  $\sigma_{st}$  denotes the number of all shortest paths from  $s$  to  $t$  and  $\sigma_{st}(T)$  is the number of shortest paths from  $s$  to  $t$  passing through all nodes from  $T$  (if  $s \in T$  or  $t \in T$  then  $\sigma_{st}(T) = 0$ ) such that the order of visited nodes is maintained (for path  $p$  we have  $T \preceq p$ ).

Similarly to other game-theoretic extension of classic centrality measures recently considered in the literature (see Section 3.2), the above function, that assigns values to all ordered groups (coalitions) of nodes, can be considered to be a value function, i.e., in our case,

the generalized characteristic function. Now, the NR and SB values computed for this function become the *generalized game-theoretic betweenness centrality*.<sup>2</sup>

**Definition 49 (Generalized representation function).** A generalized representation function is a function  $\psi^g : \mathcal{G} \rightarrow \mathcal{V}_G^g$  that maps every graph  $G = (V, E)$  onto a generalized cooperative game  $(NA, \nu_G^g) \in \mathcal{V}_G^g$  with  $A = V$ .

Now, solution concepts like Nowak-Radzick and Sánchez-Bergantiños values can be used in the network setting by applying them to the characteristic function that a network represents.

**Definition 50 (Generalized Game-theoretic centrality measure).** Formally, we define a generalized game-theoretic centrality measure as a pair  $(\psi^g, \phi)$  consisting of a representation function  $\psi^g$  and a solution concept  $\phi$ .

**Definition 51 (Generalized Game-theoretic Betweenness Centrality).** The generalized game-theoretic betweenness centrality is a pair  $(\psi_B^g, \phi)$ , where  $\psi_B^g(G) = c_{pb}$  is the path betweenness centrality, the value of each node  $v \in V$  is given by  $\phi_v(c_{pb})$ , and  $\phi$  is the Nowak-Radzick value ( $\phi^{NR}$ ) or Sánchez-Bergantiños value ( $\phi^{SB}$ ).

In the previous chapters we have shown that some game-theoretic centrality measures can be computed in polynomial time in the size of the network, i.e., in the number of nodes and edges. While in the literature there is no such a positive result for the generalized game-theoretic betweenness centrality, we will now show that it is possible to compute it in time polynomial in the number of shortest paths in the network and with the help of our generalized read-once MC-Nets.

In particular, given a network  $G = (V, E)$ , Algorithm 24 builds, for each shortest path in the network, a corresponding generalized read-once rule. More specifically, after initialization, we iterate over all shortest paths in the network. For each shortest path, in line 7, we initialize the formula  $\mathcal{F}$  with ordered atomic formula corresponding to the sequence of all the nodes on the path. Next, in line 8, we add a conjunction that ensures that no other nodes can appear in an ordered coalition that satisfies our formula.

The size of the set  $\mathcal{GR}$  constructed by this algorithm equals to the number of shortest paths in the graph and can be exponential in the number of vertices. However, in many graphs we have  $|\mathcal{GR}| \ll \mathcal{T}$ , which makes this algorithm much more efficient than computing this measure directly using equation (8.3) or equation (8.6) for  $\phi^{NR}$  and  $\phi^{SB}$ , respectively.

---

<sup>2</sup> We note that the game-theoretic betweenness centrality was introduced in Chapter 5 but this measure does not account for the order of players.

---

**Algorithm 24:** Computing the generalized game-theoretic betweenness centrality

---

**Input:** Graph  $G = (V, E)$ , path betweenness centrality  $c_{pb}$ , game solution  $\phi$

**Output:**  $\phi_v$  for all  $v \in V$

```
1 for  $v \in V$  do  $\phi_v \leftarrow 0$ ;  $\mathcal{GR} \leftarrow \emptyset$ ; // initialize the set of generalized read-once formulas
2 for  $v \in V$  do
3   for  $u \in V$  do
4      $SP \leftarrow \text{shortestPaths}(v, u)$ ; // compute using a BFS algorithm all shortest paths
        from  $v$  to  $u$ 
5     for  $(v, \dots, u) \in SP$  do
6        $\mathcal{F} \leftarrow \langle v, \dots, u \rangle$ ;
7        $\mathcal{F} \leftarrow \mathcal{F} \wedge \neg\{V \setminus X_{\mathcal{F}}\}$ ;
8        $V \leftarrow c_{pb}((v, \dots, u))$ ;
9        $\mathcal{GR} \leftarrow \mathcal{GR} \cup (\mathcal{F} \rightarrow V)$ ;
10 if  $\phi == \phi^{NR}$  then
11    $\text{COMP\_NR}(\mathcal{GR})$ ; // use Algorithm 22
12 else if  $\phi == \phi^{SB}$  then
13    $\text{COMP\_SB}(\mathcal{GR})$ ; // use Algorithm 23
```

---

## 8.4. Conclusions

Generalized characteristic function games are attracting increasing interest in the literature due to their manifold potential applications. In this chapter we considered both the representation and computational aspects of these games. In particular, building upon the state-of-the-art read-once MC-Nets for the Shapley value by Elkind et al. [2009a], we proposed the first computationally-convenient formalism to model games with ordered coalitions. Moreover, our Generalized Marginal-Contribution Nets representation allows to compute effectively the Nowak-Radzik and Sánchez-Bergantiños values. Finally, we proved the usefulness of such representation by constructing efficient algorithm to compute generalized game-theoretic betweenness centrality.

## 9. Summary and Potential Extensions

In this dissertation, we developed 10 and analyzed additional 5 efficient algorithms computing different coalitional game solutions on graphs (the most important of them were presented in Table 9.1). These positive results in the domain of networks involved three fundamental centrality metrics: degree, closeness and betweenness. The game-theoretic solution concepts, such as the Shapley value, or Nowak-Radzik value, computed for these three measures gave us the new class of centrality metrics—the game-theoretic centrality measures.

In Chapter 4, two algorithms were introduced: Algorithm 1 for Shapley value-based degree centrality and Algorithm 4 for Shapley value-based closeness centrality. Both algorithms can be easily applied to weighted and unweighted graphs. In this chapter we also presented three generalizations of these two measures. Additionally, Algorithm 1 was an inspiration for Szczepański et al. [2015a], who computed efficiently on networks Interaction Index based on degree centrality.

In the next Chapter 5 four algorithms were introduced to compute different version of game-theoretic betweenness centrality. Algorithm 12 and Algorithm 15 compute Shapley value-based betweenness centrality for weighted and unweighted networks, respectively. The another two Algorithms 13 for unweighted and Algorithms 16 for weighted graphs are an implementation of Semivalue-based betweenness centrality.

Owen value-based degree centrality was introduced in Chapter 6 together with two polynomial time algorithms. Firstly, in Algorithm 17 the general method for computing Coalitional Semivalue-based centrality measure was introduced. Secondly, in Algorithm 18 we introduced the faster method to compute Owen value-based degree centrality, which was an optimization of Algorithm 17 to calculate this particular Coalitional Semivalue.

The contribution of this dissertation not only consists of positive results. In particular, in Chapter 7 it was proven that computing Shapley value for connectivity games is  $\#P$ -hard. However, against this negative results the Algorithm 20 was developed to compute Shapley value much faster than the brute force approach for connectivity games. This algorithm allows us to study much bigger real-life networks, and also was an inspiration for developing even faster methods to analyze connectivity games and graph-restricted games [Skibski et al., 2014, Michalak et al., 2015a].

In the last Chapter 8 the new representation of generalized coalitional games was introduced. Importantly, we developed two algorithms that in time polynomial in the



size of the representation computes the Nowak-Radzick value (Algorithms 22) and the Sánchez-Bergantiños value (Algorithms 23). Based on this representation the Algorithm 24 was introduced to compute Nowak-Radzick value-based and Sánchez-Bergantiños value-based betweenness centrality.

Apart from computational results, it was experimentally shown that game-theoretic centrality measures were useful to analyze scale-free and real-life networks. In Chapter 5, the extensive simulations showed that game-theoretic betweenness centrality was a better measure to protect network against simultaneous nodes failures than the standard betweenness centrality.

The experiments performed on scientific citation network in Chapter 6 suggested that Owen value-based weighted degree centrality could be successfully used to evaluate the significance of publications, while taking into account the importance of journals and conference venues. Speaking more generally, the importance of a single node stems not only from its individual power, but also from the power of community it belongs to. Owen value-based centrality measure is the first measure in literature that takes this fact into account.

In Chapter 7 a terrorist network was analyzed with a centrality build upon connectivity games. In particular, our algorithm proposed in this chapter allowed to analyze real-life WTC 9/11 terrorist network with 36 members and 125 identified connections.

There are a number of directions for future work. Table 9.1 presents our main computational results vs. the gaps in the literature. For instance it is still an open question if it is possible to compute Owen value-based betweenness centrality in polynomial time? Such measure would find interesting applications to analyze financial institution networks, in which many institutions form strictly collaborating conglomerates. As can be seen from Table 9.1 there are more open questions regarding computation aspects of the game-theoretic centrality measures.

In this dissertation we analyzed three types of coalitional games: the standard games, games with coalitional structure, and generalized coalitional games. Nevertheless, there are still other classes of coalitional games, such as games with either positive or negative externalities [Yi, 1997, Michalak et al., 2009], that have been extensively studied in game theory and that may yield interesting results when applied to network centrality.

It is also interesting to analyse the properties of game-theoretic network centralities constructed on solution concepts from cooperative game theory other than the Shapley value, or Semivalues. In particular, more general solution concepts such as the core [Osborne and Rubinstein, 1994] or the nucleolus [Schmeidler, 1969] could be also used as a centrality metric.

There are several other potential directions for future work. Firstly, it would be interesting to extend the algorithms proposed in this thesis to streaming graphs [Green et al., 2012]. The challenge here is to efficiently re-compute the centralities and update the ranking of nodes after each change in the structure of the network, i.e., after adding/removing some nodes and/or edges. Secondly, we are keen to develop polynomial-time algorithms for the most specific

Standard centrality	degree centrality	closeness centrality	betweenness centrality
Shapley value-based centrality	Algorithm 1	Algorithm 4	Algorithm 12
Semivalue-based centrality	[Szczepański et al., 2015b]		Algorithm 13
Owen value-based centrality	Algorithm 18	[Tarkowski et al., 2016]	<b>unknown</b>
Coalitional Semivalue-based centrality	Algorithm 17	[Tarkowski et al., 2016]	<b>unknown</b>
Nowak-Radzik-based centrality	<b>unknown</b>	<b>unknown</b>	Algorithm 24 <sup>(*)</sup>
Sánchez-Bergantiños-based centrality	<b>unknown</b>	<b>unknown</b>	Algorithm 24 <sup>(*)</sup>

(\*) This algorithm does not run in polynomial time.

Table 9.1: Summary of the 7 algorithms for game-theoretic centrality measures presented in this dissertation. The above table shows also gaps in the literature regarding other computational results.

variation of centralities analyzed in this dissertation, such as routing betweenness centrality [Dolev et al., 2010] built upon betweenness centrality. Finally, we would like to identify various axiomatic systems that each uniquely characterize our centrality measure. Such an axiomatic approach is common in the literature on cooperative game theory, and we believe it would be interesting to follow a similar approach to analyse centrality measures in general, and ours in particular.

Furthermore, our contribution on representation can be extended in various directions, two of them seem particularly interesting. On the one hand, it would be interesting to consider representation formalism other than MC-Nets, such as Algebraic Decision Diagrams [Aadithya et al., 2011, Sakurai et al., 2011], to model generalized characteristic function games. On the other hand, it is worth analyzing if our representation can produce algorithms that compute generalized game-theoretic centrality measures build upon degree and closeness centralities.

Finally, it would be worth developing a more formal and general approach that would allow us to construct coalitional games defined over networks that correspond to other known centrality metrics or even entire families of them. Such an approach would involve developing a group centrality first and then building a characteristic function of a coalitional game upon it. Of course, while developing new centrality metrics based on coalitional games, one should keep in mind computational properties of the proposed solutions. Although we were able to obtain satisfactory computational results for a number of games, the computation of the game-theoretic network centrality may become much more challenging for more complex definitions of the characteristic function.

# Appendix A. Main Notation Used in the Dissertation

## Graphs/Networks

---

$V = \{v_1, v_2, \dots, v_{ V }\}$	The set of nodes.
$E = \{e_1, e_2, \dots, e_{ E }\}$	The set of undirected (two-elements sets) or directed (two-elements ordered pairs) edges $e = (v_i, v_j)$ .
$G = (V, E)$	The simple graph, where $E$ is set of undirected edges.
$G = (V, E)$	The directed graph, where $E$ is set of directed edges.
$G = (V, E, \lambda)$	The weighted graph with weight function: $\lambda(v, u) : E \rightarrow \mathbb{R}$ .
$G = (V, E, \gamma)$	The weighted graph with weights on nodes: $\gamma : V \rightarrow \mathbb{R}$ .
$N(v)$	The neighbors of the node $v$ , that is: $\{u : (v, u) \in E\}$ .
$\deg(v)$	Degree of the node $v$ , $\deg(v) =  N(v) $ .
$N(C)$	The neighbors of the set of nodes $C \subseteq V$ , i.e. $\bigcup_{v_i \in C} N(v_i) \setminus C$ .
$E(C)$	The set of edges within set of nodes $C$ : $E(C) = \{(u, v) \in E : u, v \in C\}$ .
$\deg(C)$	The degree of the set $C$ , i.e. $\deg(C) =  N(C) $ .
$\text{dist}(v, u)$	The distance between two nodes: $v$ and $u$ .
$\text{dist}(C, v)$	The distance between two nodes: $v$ and set of nodes $C$ : $\text{dist}(C, v) = \min_{u \in C} \text{dist}(u, v)$ .
$c_d, c_b, c_c : V \rightarrow \mathbb{R}$	The degree, betweenness and closeness centrality, respectively.
$c_{gd}, c_{gb}, c_{gc} : 2^V \rightarrow \mathbb{R}$	The group degree, group betweenness and group closeness centrality, respectively.
$C \subseteq V$	A community $C$ .
$CS = \{C_1, C_2, \dots, C_{ CS }\}$	The community structure.
$\sigma_{st}$	The number of shortest paths between $s$ and $t$ .
$\sigma_{st}(v)$	The number of shortest paths between $s$ and $t$ passing through the node $v$ .
$\sigma_{st}(C)$	The number of shortest paths between $s$ and $t$ passing through at least one node from $C$ .

## Coalitional Games

---

$A = \{a_1, a_2, \dots, a_{ A }\}$	The set of players/agents.
$\nu : 2^A \rightarrow \mathbb{R}$	The characteristic function.
$g = (A, \nu)$	A coalitional game $g$ in a characteristic function form.
$C \subseteq A$	A coalition $C$ .
$\Pi(C)$	The set of all orders (permutations) of players in $C$ .
$\text{MC}(C, a_i)$	The marginal contribution of player $a_i$ to coalition $C$ .
$\phi_i(A, \nu)$	A solution concept for a player $a_i$ in coalitional game $(A, \nu)$ .
$\phi_i(A)$	Shorter notation for $\phi_i(A, \nu)$ .
$\phi_i^{SV}(A, \nu)$	The Shapley value for a player $a_i$ in coalitional game $(A, \nu)$ .
$\phi_i^{SEMI}(A, \nu)$	The Semivalue for a player $a_i$ in coalitional game $(A, \nu)$ .
$\psi : \mathcal{G} \rightarrow \mathcal{V}$	The representation function ( $\mathcal{G}$ is the set of all graphs, and $\mathcal{V}$ the set of all coalitional games).
$(A, \mathcal{R})$	The coalitional game in the form of Marginal Contribution Network.
$\delta_j^i$	The difference between $\phi_i(N) - \phi_i(N_{-j})$ and $\phi_j(N) - \phi_j(N_{-i})$ .

## Coalitional Games with Coalitional Structure

---

$CS = \{C_1, C_2, \dots, C_{ CS }\}$	The coalitional structure.
$\phi_i^{OV}(A, \nu, CS)$	The Owen value for a player $a_i$ in coalitional game $(A, \nu, CS)$ with coalitional structure $CS$ .
$\phi_i^{SEMI}(A, \nu, CS)$	The Coalitional Semivalue for a player $a_i$ in coalitional game $(A, \nu, CS)$ with coalitional structure $CS$ .

#### Generalized Coalitional Games

---

$T$	An ordered coalition.
$\mathcal{T} = \bigcup_{C \in 2^N} \Pi(C)$	The set of all ordered coalitions.
$\nu^g : \mathcal{T} \rightarrow \mathbb{R}$	A generalized characteristic function.
$g^g = (A, \nu^g)$	A generalized coalitional game in characteristic function form.
$(T, T')^k$	The ordered coalition that results from inserting $T'$ at the $k^{th}$ position in $T$ .
$(T, a_i)$	the ordered coalition that results from inserting $a_i$ at the end of $T$ .
$(A, \nu^g)$	A coalitional game in a generalized characteristic function form.
$T(a_i)$	An ordered coalition made of all predecessors of player $a_i$ in $T$ .
$\phi_i^{SB}(N, \nu^g)$	The Sánchez-Bergantiños value of player $a_i$ in game $(N, \nu^g)$ .
$\phi_i^{NR}(N, \nu^g)$	The Nowak-Radzick value of player $a_i$ in game $(N, \nu^g)$ .
$v_g^{NR}(C)$	The Nowak-Radzick value of coalition $C$ (i.e., $\frac{1}{ C !} \cdot \sum_{T \in \Pi(C)} \nu^g(T)$ ).
$MC^{NR}(T, i)$	$\nu^g((T, a_i)) - \nu^g(T)$ .
$MC^{SB}(T, i)$	$\frac{1}{( T +1)} \sum_{l=1}^{ T +1} [\nu((T, a_i)^l) - \nu(T)]$ .
$(N, \mathcal{GR})$	generalized coalitional game represented by generalized MC-net

#### Game-theoretic Network Centralities

---

$\nu_G : 2^V \rightarrow \mathbb{R}$	The characteristic function defined on graph.
$g = (N, \nu_G)$	A coalitional game $g$ in a characteristic function form defined on graph.

#### Other

---

$\mathbb{E}[\cdot]$	The expectation operator.
$(z \wedge \neg z') \rightarrow V$	Basic MC-Nets $z$ positive literals, $z'$ positive literals,.
$\mathcal{F}$	Formula
$\mathcal{F}_{BAF} = \{a_m, \dots, a_n\}$	basic atomic formula (BAF)
$\mathcal{F}_{OAF} = \langle a_i, \dots, a_j \rangle$	ordered atomic formula (OAF)
$\langle a_i^1, \dots, a_j^f \rangle$	upper indices indicate positions in formula/ordered coalition
$\mathcal{F}_1 \oplus \mathcal{F}_2$	exclusive disjunction
$T \models \mathcal{F}$	ordered coalition $T$ meets or satisfies formula $\mathcal{F}$
$T \not\models \mathcal{F}$	ordered coalition $T$ does not meet or satisfy formula $\mathcal{F}$
$X_{\mathcal{F}}$	set of players occurring in $\mathcal{F}$
$\mathcal{T}_{\mathcal{F}}$	$\bigcup \{\Pi(C) : C \subseteq X_{\mathcal{F}}\}$
$T_1 \times T_2$	conflation of two ordered coalitions
$A_{k,i}(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, a_i \notin T, T \not\models \mathcal{F}, (T, a_i) \models \mathcal{F}\} $
$B_{k,i}(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, a_i \notin T, T \models \mathcal{F}, (T, a_i) \not\models \mathcal{F}\} $
$T_k(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, T \models \mathcal{F}\} $
$F_k(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, T \not\models \mathcal{F}\} $
$T_k(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, a_i \notin T, T \not\models \mathcal{F}, (T, a_i)^l \models \mathcal{F}\} $
$B_{k,i,l}(\mathcal{F})$	$ \{T \in \mathcal{T}_{\mathcal{F}} :  T  = k, a_i \notin T, T \models \mathcal{F}, (T, a_i)^l \not\models \mathcal{F}\} $

# Appendix B. The Recursive Functions

---

## Algorithm 25: $Sh^{NR}(\mathcal{F})$

---

**Input:** Formula  $\mathcal{F}$   
**Output:** Quantities  $(A, B, T, F)$

```

begin
  if  $\mathcal{F} = \{a_1, \dots, a_r\}$  then
     $(A, B, T, F) \leftarrow (0, 0, 0, 0)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}$  do
        if  $k = r - 1$  then  $A_{k,i} \leftarrow (r - 1)!$ 
      end
      if  $k = r$  then  $T_k \leftarrow r!$   $F_k \leftarrow k! - T_k$ ;
    end
  end
  else if  $\mathcal{F} = \langle a_1, \dots, a_r \rangle$  then
     $(A, B, T, F) \leftarrow (0, 0, 0, 0)$ ;
    for  $k \in \{0, \dots, r\}$  do
      if  $k = r - 1$  then  $A_{k,r} \leftarrow 1$  if  $k = r$  then  $T_k \leftarrow 1$   $F_k \leftarrow k! - T_k$ ;
    end
  end
  else if  $\mathcal{F} = \neg \mathcal{F}_1$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{NR}(\mathcal{F}_1)$ ;
     $(A, B, T, F) \leftarrow (B^1, A^1, F^1, T^1)$ ;
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \wedge \mathcal{F}_2$  then // We assume that  $\forall_{i \in X_{\mathcal{F}}} i \in \mathcal{F}_1$ , the second situation is analogous.
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{NR}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{NR}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}$  do
         $A_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} A_{s,i}^1 T_{k-s}^2$ 
         $B_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} B_{s,i}^1 T_{k-s}^2$ 
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} T_s^1 T_{k-s}^2$ 
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (F_s^1 F_{k-s}^2 + F_s^1 T_{k-s}^2 + T_s^1 F_{k-s}^2)$ 
    end
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \vee \mathcal{F}_2$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{NR}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{NR}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}$  do
         $A_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} A_{s,i}^1 F_{k-s}^2$ 
         $B_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} B_{s,i}^1 F_{k-s}^2$ 
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 T_{k-s}^2 + F_s^1 T_{k-s}^2 + T_s^1 F_{k-s}^2)$ 
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} F_s^1 F_{k-s}^2$ 
    end
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \oplus \mathcal{F}_2$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{NR}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{NR}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}$  do
         $A_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} (B_{s,i}^1 T_{k-s}^2 + A_{s,i}^1 F_{k-s}^2)$ 
         $B_{k,i} \leftarrow \sum_{s=0}^k \binom{k}{k-s} (B_{s,i}^1 F_{k-s}^2 + A_{s,i}^1 T_{k-s}^2)$ 
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 F_{k-s}^2 + F_s^1 T_{k-s}^2)$ 
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 T_{k-s}^2 + F_s^1 F_{k-s}^2)$ 
    end
  end
end

```

---

---

**Algorithm 26:**  $Sh^{SB}(\mathcal{F})$ 


---

**Input:** Formula  $\mathcal{F}$

**Output:** Quantities  $(A, B, T, F)$

```

begin
  if  $\mathcal{F} = \{a_1, \dots, a_r\}$  then
     $(A, B, T, F) \leftarrow (0, 0, 0, 0)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}, l \in \{1, \dots, r\}$  do
        if  $k = r - 1$  then  $A_{k,i,l} \leftarrow (r - 1)!$ 
      end
      if  $k = r$  then  $T_k \leftarrow r! \quad F_k \leftarrow k! - T_k$ ;
    end
  end
  else if  $\mathcal{F} = \langle a_1, \dots, a_r \rangle$  then
     $(A, B, T, F) \leftarrow (0, 0, 0, 0)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}$  do
        if  $k = r - 1$  then  $A_{k,i,i} \leftarrow 1$ 
      end
      if  $k = r$  then  $T_k \leftarrow 1 \quad F_k \leftarrow k! - T_k$ ;
    end
  end
  else if  $\mathcal{F} = \neg \mathcal{F}_1$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{SB}(\mathcal{F}_1)$ ;
     $(A, B, T, F) \leftarrow (B^1, A^1, F^1, T^1)$ ;
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \wedge \mathcal{F}_2$  then // We assume that  $\forall_{i \in X_{\mathcal{F}}} i \in \mathcal{F}_1$ , the second situation is analogous.
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{SB}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{SB}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}, l \in \{1, \dots, r\}$  do
         $A_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} A_{s,i,m}^1 T_{k-s}^2$ ;
         $B_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} B_{s,i,m}^1 T_{k-s}^2$ ;
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} T_s^1 T_{k-s}^2$ ;
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (F_s^1 F_{k-s}^2 + F_s^1 T_{k-s}^2 + T_s^1 F_{k-s}^2)$ ;
    end
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \vee \mathcal{F}_2$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{SB}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{SB}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}, l \in \{1, \dots, r\}$  do
         $A_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} A_{s,i,m}^1 F_{k-s}^2$ ;
         $B_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} B_{s,i,m}^1 F_{k-s}^2$ ;
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 T_{k-s}^2 + F_s^1 T_{k-s}^2 + T_s^1 F_{k-s}^2)$ 
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} F_s^1 F_{k-s}^2$ 
    end
  end
  else if  $\mathcal{F} = \mathcal{F}_1 \oplus \mathcal{F}_2$  then
     $(A^1, B^1, T^1, F^1) \leftarrow Sh^{SB}(\mathcal{F}_1)$ ;
     $(A^2, B^2, T^2, F^2) \leftarrow Sh^{SB}(\mathcal{F}_2)$ ;
    for  $k \in \{0, \dots, r\}$  do
      for  $i \in X_{\mathcal{F}}, l \in \{1, \dots, r\}$  do
         $A_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} (B_{s,i,m}^1 T_{k-s}^2 + A_{s,i,m}^1 F_{k-s}^2)$ ;
         $B_{k,i,l} \leftarrow \sum_{s=0}^k \sum_{m=1}^{s+1} \binom{l-1}{l-m} \binom{k-l+1}{k-l+m-s} (B_{s,i,m}^1 F_{k-s}^2 + A_{s,i,m}^1 T_{k-s}^2)$ ;
      end
       $T_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 F_{k-s}^2 + F_s^1 T_{k-s}^2)$ ;
       $F_k \leftarrow \sum_{s=0}^k \binom{k}{k-s} (T_s^1 T_{k-s}^2 + F_s^1 F_{k-s}^2)$ ;
    end
  end
end

```

---

# Bibliography

- K. V. Aadithya, T. P. Michalak, and N. R. Jennings. Representation of coalitional games with algebraic decision diagrams. In *Proceedings of 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1121–1122, 2011.
- K. Adamczewski, S. Matejczyk, and T. P. Michalak. How good is the shapley value-based approach to the influence maximization problem? In *ECAI'14 - 21st European Conference on Artificial Intelligence*, pages 955–956, 2014.
- R. Albert, H. Jeong, and A.L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.
- H. Alsuwaiyel. *Algorithms: Design Techniques and Analysis*. Lecture Notes Series on Computing. World Scientific, 1999.
- R. Amer and J. M. Gimenez. A connectivity game for graphs. *Math. Methods of OR*, 60: 453–470, 2004.
- R. Amer, F. Carreras, and J. M. Giménez. The modified banzhaf value for games with coalition structure: an axiomatic characterization. *Mathematical Social Sciences*, 43(1):45–54, 2002.
- R. Amer, J. Miguel G., and A. Magaña. Accessibility in oriented networks. *European Journal of Operational Research*, 180(2):700 – 712, 2007. ISSN 0377-2217.
- J.M. Anthonisse. The rush in a directed graph., 1971. Tech. Rep. BN 9/71.
- H. Aziz and B. de Keijzer. Complexity of coalition structure generation. In *AAMAS '11: Tenth International Conference on Autonomous Agents and Multi-Agent Systems*, pages 191–198, 2011.
- H. Aziz, O. Lachish, M. Paterson, and R. Savani. Wiretapping a hidden network. In *WINE '09: Proceedings of the the 5th Workshop on Internet & Network Economics*, pages 438–446, 2009.
- R. Bacher. Determinants of matrices related to the pascal triangle. *JTNB*, 14(1):19–41, 2002.
- Y. Bachrach and J. S. Rosenschein. Coalitional skill games. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2, AAMAS '08*, pages 1023–1030, 2008.
- Y. Bachrach and J.S. Rosenschein. Power in threshold network flow games. *Autonomous Agents and Multi-Agent Systems*, 18(1):106–132, 2009.
- Y. Bachrach, E. Markakis, A. D. Procaccia, J. S. Rosenschein, and A. Saberi. Approximating power indices. In *AAMAS '08: Proceedings of the 7th International Joint Conference on*

- Autonomous Agents and Multi-Agent Systems*, pages 943–950, 2008.
- M. O. Ball, B. L. Golden, and R. V. Vohra. Finding the most vital arcs in a network. *Oper. Res. Lett.*, 8(2):73–76, 1989.
- J. F. Banzhaf. Weighted Voting Doesn’t Work: A Mathematical Analysis. *Rutgers Law Review*, 19:317–343, 1965.
- A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means*. Plume, 2003.
- A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change in informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.
- P. Boldi and S. Vigna. Axioms for centrality. *CoRR*, abs/1308.2140, 2013.
- E. Bompard, Di Wu, and Fei Xue. The concept of betweenness in the analysis of power grid vulnerability. In *Complexity in Engineering, 2010. COMPENG ’10.*, pages 52–54, 2010.
- P. Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- P. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.
- R. B. Boppana and M. Sipser. The complexity of finite functions. In *Handbook of Theoretical Computer Science, Volume A: Algorithms and Complexity (A)*, pages 757–804. 1990.
- S. Borgatti, C. Jones, and M. Everett. Network Measures of Social Capital. *Connections*, 21(2):27–36, 1998.
- S. P. Borgatti and M. G. Everett. A Graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.
- U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
- U. Brandes. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks*, 30(2):136–145, 2008.
- U. Brandes and T. Erlebach. *Network Analysis: Methodological Foundations*. Lecture notes in computer science: Tutorial. Springer, 2005. ISBN 9783540249795.
- S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.
- R. S. Burt. *Structural holes: The social structure of competition*. Harvard University Press, 1992.
- D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network Robustness and Fragility: Percolation on Random Graphs. *Physical Review Letters*, 85(25):5468–5471, 2000.



- K.M. Carley, J. Reminga, and N. Kamneva. Destabilizing terrorist networks. In *NAACSOS*, 2003.
- F. Carreras and M. Puente. Symmetric coalitional binomial semivalues. *Group Decision and Negotiation*, 21(5):637–662, 2012. ISSN 0926-2644.
- A. Casajus. Marginality, differential marginality, and the banzhaf value. *Theory and Decision*, 71(3):365–372, 2011.
- J. Castro, D. Gomez, and J. Tejada. Polynomial calculation of the shapley value based on sampling. *Computers & Operations Research*, 36(5):1726–1730, 2009.
- J.-F. Caulier. A note on the monotonicity and superadditivity of TU cooperative games. April 2009.
- G. Chalkiadakis, E. Elkind, and M. Wooldridge. *Computational Aspects of Cooperative Game Theory*. Synthesis Lec. on AI and ML. 2011.
- V. Conitzer and T. Sandholm. Complexity of constructing solutions in the core based on synergies among coalitions. *Artificial Intelligence*, pages 170(6):607–619, 2006.
- H.W. Corley and D. Y. Sha. Most vital links and nodes in weighted networks. *Operations Research Letters*, 1(4):157–160, 1982.
- T.H. Cormen. *Introduction to algorithms*. MIT Press, 2001.
- K. J. Cormican, D. P. Morton, and R. K. Wood. Stochastic Network Interdiction. *Operations Research*, 46(2), 1998.
- M. del Pozo, C. Manuel, E. González-Arangüena, and G. Owen. Centrality in directed social networks. a game theoretic approach. *Social Networks*, 33(3):191 – 200, 2011. ISSN 0378-8733.
- X. Deng and C. Papadimitriou. On the complexity of cooperative solution concepts. *Mathematics of Operations Research*, 19(2):257–266, 1994.
- S. Dolev, Y. Elovici, and R. Puzis. Routing betweenness centrality. *Journal of ACM*, 57(4): 25:1–25:27, May 2010.
- P. Dubey, A. Neyman, and R. J. Weber. Value Theory Without Efficiency. *Mathematics of Operations Research*, 6:122–128, 1981.
- E. Elkind, L. Ann Goldberg, P. W. Goldberg, and M. Wooldridge. A tractable and expressive class of marginal contribution nets and its applications. *Mathematical Logic Quarterly*, 55 (4):362 – 376, 2009a.
- E. Elkind, L.A. Goldberg, P.W. Goldberg, and M. Wooldridge. A tractable and expressive class of marginal contribution nets and its applications. *Mathematical Logic Quarterly*, 55(4): 362–376, 2009b.
- European Central Bank. Recent advances in modelling systemic risk using network analysis. Technical report, 2010.

- M. G. Everett and S. P. Borgatti. The centrality of groups and classes . *Journal of Mathematical Sociology*, 23(3):181–201, 1999.
- J.D. Farley. Breaking al qaeda cells: a mathematical analysis of counterterrorism operations. *Studies in Conflict and Terrorism*, 31:126–137, 2003.
- S. S. Fatima, M. Wooldridge, and N.R. Jennings. A randomized method for the shapley value for the voting game. In *AAMAS '07: Proceedings of the 11th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 955–962, 2007.
- S. S. Fatima, M. Wooldridge, and N.R. Jennings. A linear approximation method for the shapley value. *Artificial Intellilgence*, 172(14):1673–1699, September 2008.
- Financial Stability Board. Reducing the moral hazard posed by systemically important financial institutions. Technical report, 2010.
- S Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75 – 174, 2010.
- L.C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3): 215–239, 1979.
- X. Freixas, B. Parigi, and J. Ch. Rochet. Systemic risk, interbank relations and liquidity provision by the central bank. *Journal of Money Credit and Banking*, (32), 2000.
- N.E. Friedkin. Theoretical foundations for centrality measures. *American Journal of Sociology*, 96(6):1478—1504, 1991.
- S. Gabrieli. Too-interconnected versus too-big-to-fail: Banks’ network centrality and overnight interest rates. Working paper series, Social Science Research Network, February 2011.
- M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12): 7821–7826, June 2002.
- D. Gómez, E. González-Arangüena, C. Manuel, G. Owen, M. Del Pozo, and J. Tejada. Centrality and power in social networks: A game theoretic approach. *Mathematical Social Sciences*, 46(1):27–54, 2003.
- A. Goyal, F. Bonchi, and L. V.S. Lakshmanan. Learning influence probabilities in social networks. In *WSDM '10: Proceedings of the 3rd ACM international conference on Web search and data mining*, pages 241–250, 2010.
- M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83 (6):1420–1443, 1978.
- G. Greco, E. Malizia, L. Palopoli, and F. Scarcello. On the complexity of compact coalitional games. In *IJCAI '09: Proceedings of the Twenty First International Joint Conference on Artificial Intelligence*, pages 147–152, 2009.
- O. Green, R. McColl, and D. A. Bader. A fast algorithm for streaming betweenness centrality. In *SocialCom/PASSAT*, pages 11–20, 2012.

- B. Grofman and Guillermo Owen. A game-theoretic approach to measuring centrality in social networks. *Social Networks*, 4:213–224, 1982.
- R. Guimerà, L. Danon, A. Diaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(065103), 2003.
- G.A. Henke. How Terrorist Groups Survive: A Dynamic Network Analysis Approach to the Resilience of Terrorist Organizations, 2009.
- P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Physical Review E*, 65(5):056109, 2002.
- B. Husslage, R. H. A. Lindelauf, and H. Hamers. Leaderless Covert Networks: A Quantitative Approach, 2012.
- I2. Analyst’s Notebook 8, Social Network Analysis Whitepaper. [www.i2group.com](http://www.i2group.com), 2010.
- S. Jeong and Y. Shoham. Marginal contribution nets: a compact representation scheme for coalitional games. In *ACM EC ’05: Proceedings of the Sixth ACM Conference on Electronic Commerce*, pages 193–202, 2005.
- H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, May 2001.
- D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *KDD ’03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.
- D. Kempe, J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. *Automata, Languages and Programming*, 3580:99, 2005.
- B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer, 2005. ISBN 9783540256847.
- D. Koschützki, K.A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski. Centrality indices. *Network analysis*, volume 3418 of *Lecture Notes in Computer Science*, pages 16–61. Springer, 2005.
- V.E. Krebs. Mapping networks of terrorist cells. *Connections*, 24:43–52, 2002.
- L. Laeven and F. Valencia. *Resolution of banking crises: The good, the bad, and the ugly*. International Monetary Fund, 2010.
- M.-J. Lee, J. Lee, J. Y. Park, R. H. Choi, and C.-W. Chung. Qube: a quick algorithm for updating betweenness centrality. In Alain Mille, Fabien L. Gandon, Jacques Misselis, Michael Rabinovich, and Steffen Staab, editors, *WWW*, pages 351–360. ACM, 2012. ISBN 978-1-4503-1229-5.
- R. Lindelauf, H. Hamers, and B. Husslage. Cooperative game theoretic centrality analysis of terrorist networks: The cases of jemaah islamiyah and al qaeda. *European Journal of Operational Research*, 229(1):230–238, 2013.

- R.H.A. Lindelauf, P. Borm, and H.J.M. Hamers. The influence of secrecy on the communication structure of covert networks. *Social Networks*, 31:126–137, 2009a.
- R.H.A. Lindelauf, P. Borm, and H.J.M. Hamers. *On Heterogeneous Covert Networks*. Springer, 2009b.
- H. Liu, Z. Hu, H. Haddadi, and H. Tian. Hidden link prediction based on node centrality and weak ties. *EPL (Europhysics Letters)*, 101(1), 2013.
- K. Malik, A. K. Mittal, and S. K. Gupta. The k most vital arcs in the shortest path problem. *Oper. Res. Lett.*, 8(4), 1989.
- I. Mann and L. S. Shapley. *Values of Large Games, IV : Evaluating the Electoral College by Montecarlo Techniques*. CA : RAND Corporation, Santa Monica, 1960.
- I. Mann and L.S. Shapley. *Values of Large Games. Vi: Evaluating the Electoral College Exactly*. Rand Corporation Memorandum RM-3158-PR. Defense Technical Information Center, 1962.
- N. Memon, H. Larsen, H. Legind, D. Hicks, and N. Harkiolakis. Detecting hidden hierarchy in terrorist networks: Some case studies. In *IEEE ISI'08: Proceedings of the IEEE Workshop on Intelligence and Security Informatics (ISI)*, pages 477–489, 2008.
- P.T. Michalak, K.V. Aaditha, P.L. Szczepański, B. Ravindran, and N. R. Jennings. Efficient computation of the shapley value for game-theoretic network centrality. *Journal of Artificial Intelligence Research*, 46:607–650, 2013a.
- T. P. Michalak, T. Rahwan, J. Sroka, A. Dowell, M. J. Wooldridge, P. J. McBurney, and N. R. Jennings. On representing coalitional games with externalities. In *Proceedings of the 10th ACM Conference on Electronic Commerce, EC '09*, pages 11–20, 2009.
- T. P. Michalak, D. Marciniak, M. Szamotulski, T. Rahwan, M. Wooldridge, P. McBurney, and N. R. Jennings. A logic-based representation for coalitional games with externalities. In *AAMAS*, pages 125–132, 2010.
- T. P. Michalak, P. L. Szczepański, T. Rahwan, A. Chrobak, S. Brânzei, M. Wooldridge, and N. R. Jennings. Implementation and computation of a value for generalized characteristic function games. *ACM Transactions on Economics and Computation*, 2(4):16:1–16:35, 2014.
- T. P. Michalak, T. Rahwan, O. Skibski, and M. Wooldridge. Defeating terrorist networks with game theory. *IEEE Intelligent Systems*, 30(1):53–61, 2015a.
- T.P. Michalak, T. Rahwan, P.L. Szczepański, O. Skibski, R. Narayanam, M.J. Wooldridge, and N.R. Jennings. Computational analysis of connectivity games with applications to the investigation of terrorist networks. pages 293–301, 2013b.
- T.P. Michalak, S. Moretti, R. Namarayan, O. Skibski, P. Szczepański, T. Rahwan, and M. Wooldridge. A new approach to measure social capital using game-theoretic techniques. *ACM SIGecom Exchanges*, 14(1):95–100, 2015b.
- S. Misra, R. Barthwal, and M.S. Obaidat. Community detection in an integrated internet of things and social network architecture. In *Global Communications Conference*

- (*GLOBECOM*), 2012 *IEEE*, pages 1647–1652, 2012.
- G. Moerkotte and T. Neumann. Analysis of two existing and one new dynamic programming algorithm for generation of optimal bushy join trees without cross products. In *VLDB*, 2006.
- D. Monderer and D. Samet. Variations on the shapley value. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory with Economic Applications*, volume 3, chapter 54, pages 2055–2076. Elsevier, 1 edition, 2002.
- R.B. Myerson. Graphs and cooperation in games. *Mathematics of Operations Research*, 2(3): 225–229, 1977.
- H. Nagamochi, D. Zeng, N. Kabutoya, and T. Ibaraki. Complexity of the minimum base game on matroids. *Mathematics of Operations Research*, 22(1):146–164, 1997.
- R. Narayanam, O. Skibski, H. Lamba, and T. P. Michalak. A shapley value-based approach to determine gatekeepers in social networks with applications. In *ECAI’14: 21st European Conference on Artificial Intelligence*, pages 651–656, 2014.
- M E Newman. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582, 2006.
- M. E. J. Newman. Analysis of weighted networks. *Physical Review E*, 70(5):056131, 2004.
- M.E.J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132(1–7), 2001.
- A. S. Nowak and T. Radzik. The shapley value for n-person games in generalized characteristic function form. *Games and Economic Behavior*, 6(1):150–161, 1994.
- M. J. Osborne and A. Rubinstein. *A course in game theory*. The MIT Press, 1994.
- G. Owen. Values of games with a priori unions. In *Mathematical economics and game theory: Essays in honor of Oskar Morgenstern*, volume 141 of *Lecture Notes in Economics and Mathematical Systems*, pages 76–88. Springer, 1977.
- G. Owen. Modification of the Banzhaf-Coleman index for games with a priori unions. In *Power, Voting and Voting Power*, pages 232–238. 1982.
- R. Puzis, Y. Elovici, and S. Dolev. Finding the most prominent group in complex networks. *AI Communication*, 20(4):287–296, December 2007a. ISSN 0921-7126.
- Rami Puzis, Yuval Elovici, and Shlomi Dolev. Fast algorithm for successive computation of group betweenness centrality. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 76(5), 2007b.
- T. Rahwan, T. Michalak, M. Wooldridge, and N. R. Jennings. Anytime coalition structure generation in multi-agent systems with positive or negative externalities. *Artificial Intelligence*, 186(0):95 – 122, 2012.
- T. Rahwan, T. P. Michalak, and M. Wooldridge. A measure of synergy in coalitions. *CoRR*, abs/1404.2954, 2014.

- S. Ressler. Social network analysis as an approach to combat terrorism: past, present and future research. *Homeland Security Affairs*, 2:1–10, 2006.
- M. Sageman. *Understanding Terror Networks*. University of Pennsylvania Press, 2004.
- Y. Sakurai, S. Ueda, A. Iwasaki, S. Minato, and M. Yokoo. A compact representation scheme of coalitional games based on multi-terminal zero-suppressed binary decision diagrams. In *Proceedings of 12th International Conference on Principles of Practice in Multi-Agent Systems (AAMAS)*, pages 4–18, 2011.
- E. Sánchez and G. Bergantiños. Coalitional values and generalized characteristic functions. *Math. Methods of OR*, 49(3):413–433, 1999. ISSN 1432-2994.
- T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohme. Coalition structure generation with worst case guarantees. *Artificial Intelligence*, 111(1–2):209–238, 1999. ISSN 0004-3702.
- D. Schmeidler. The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics*, pages 1163–1170, 1969.
- D. Schultes and P. Sanders. Dynamic highway-node routing. In *SEA '07: Proceedings of 6th Workshop on Experimental and Efficient Algorithms. LNCS*, pages 66–79. Springer, 2007.
- L. S. Shapley. A value for n-person games. In H.W. Kuhn and A.W. Tucker, editors, *In Contributions to the Theory of Games, volume II*, pages 307–317. Princeton University Press, 1953.
- L. S. Shapley and M. Shubik. A Method for Evaluating the Distribution of Power in a Committee System. *The American Political Science Review*, 48(3):787–792, 1954.
- R. Shultz, R. Godson, and Q. Hanlon. *Armed Groups and Irregular Warfare: Adapting Professional Military Education*. National Strategy Information Center, USA, 2009.
- O. Skibski, T.P. Michalak, T. Rahwan, and M. Wooldridge. Algorithms for the shapley and myerson values in graph-restricted games. In *AAMAS'14: Proceedings of the 13th International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 197–204, 2014.
- O. Skibski, T.P. Michalak, Y. Sakurai, and M. Yokoo. A pseudo-polynomial algorithm for computing power indices in graph-restricted weighted voting games. In *IJCAI'15: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 631–637, 2015.
- J.C. Smith and C. Lim. Algorithms for network interdiction and fortification games. In *Pareto Optimality, Game Theory And Equilibria*, volume 17 of *Springer Optimization and Its Applications*, pages 609–644. 2008.
- H. von Stackelberg. *The Theory of the Market Economy*. William Hodge and Co., 1952.
- N.R. Suri and Y. Narahari. A Shapley Value based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering*, 99:1–18, 2010.

- P.L. Szczepański, T.P. Michalak, and T. Rahwan. A new approach to betweenness centrality based on the shapley value. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '12, pages 239–246, 2012.
- P.L. Szczepański, T.P. Michalak, and M. Wooldridge. A centrality measure for networks with community structure based on a generalization of the owen value. In *ECAI'14: Proceedings of the 21st European Conference on Artificial Intelligence*, pages 867–872, 2014.
- P.L. Szczepański, A.S. Barcz, T.P. Michalak, and T. Rahwan. The game-theoretic interaction index on social networks with applications to link prediction and community detection. In *IJCAI'15: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 638–644, 2015a.
- P.L. Szczepański, M.K. Tarkowski, T.P. Michalak, P. Harrenstein, and M. Wooldridge. Efficient computation of semivalues for game-theoretic network centrality. *AAAI '15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 1:461–469, 2015b.
- P.L. Szczepański, T.P. Michalak, and T. Rahwan. Efficient algorithms for a game-theoretic betweenness centrality. *Artificial Intelligence*, 2016:39–63, 2016.
- Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: Extraction and mining of academic social networks. In *KDD'08*, pages 990–998, 2008.
- M. Tarkowski, P.L. Szczepański, T. Rahwan, T.P. Michalak, and M. Wooldridge. Closeness centrality for networks with overlapping community structure. *AAAI '16: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, page forthcoming, 2016.
- L. Tran-Thanh, T.-D. Nguyen, T. Rahwan, A. Rogers, and N. R. Jennings. An efficient vector-based representation for coalitional games. In Francesca Rossi, editor, *IJCAI*, 2013.
- A. Valdes and K. Skinner. Probabilistic alert correlation. In *Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection*, RAID '00, pages 54–68, London, UK, UK, 2001. Springer-Verlag. ISBN 3-540-42702-3. URL <http://dl.acm.org/citation.cfm?id=645839.670734>.
- T.W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1):69–89, 1996.
- R. van den Brink and P. Borm. Digraph Competitions and Cooperative Games. *Theory and Decision*, 53(4):327–342, 2002.
- R. van den Brink and R. P. Gilles. Measuring domination in directed networks. *Social Networks*, 22(2):141–157, 2000.
- T. van Laarhoven and E. Marchiori. Robust community detection methods with resolution parameter for complex detection in protein protein interaction networks. In *Pattern Recognition in Bioinformatics*, volume 7632 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 2012.

- D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684): 440–442, 1998.
- D.J.A. Welsh. *Approximate counting*, volume 241. Cambridge University Press, 1997.
- M. Wooldridge and P. Dunne. On the computational complexity of coalitional resource games. *Artificial Intelligence*, 170(10):835–871, 2006.
- J. Xu, D. Hu, and H. Chen. Dynamics of terrorist network: Understanding the survival mechanisms of global salafi jihad. *J. of Homeland Security and Emergency Management*, 6, 2009.
- S.-S. Yi. Stable coalition structures with externalities. *Games and Economic Behavior*, 20(2): 201–237, 1997.
- H.P. Young. *The diffusion of innovations in social networks. Economy as an evolving complex system III*, pages 267–281. Oxford University Press, 2005.
- W.W. Zachary. An information flow model for conflict and fission in small groups. *J. of Anthropological Research*, 33:452–473, 1977.